

THIS WEEK

EDITORIALS

HISTORY The impact of science and technology on the First World War **p.6**

WORLD VIEW Russian reforms are making life hard a year on **p.7**

MATERIALS Self-folding grippers can grab single cells **p.9**

Academic freedom under threat

The human rights of academics in Egypt are being eroded by the military regime that has taken control of the country. The Arab Spring is on hold.

Emad Shahin, a political-sciences scholar, has been in exile since January, when the Egyptian authorities issued a warrant for his arrest. He was at a conference in the United States at the time and, fearing that he will not get a fair trial in Egypt, he has not been home since. The charges against him, which Shahin declares are “ludicrous”, include espionage and being a leading member of the Muslim Brotherhood, the party of former president Mohammed Morsi, who was ousted in a military coup last July. The new regime has declared the Muslim Brotherhood a terrorist organization.

It is true that Shahin has been vocally critical of the new regime. As an internationally renowned professor of public policy and administration at the American University in Cairo, he was also critical of the repressive 30-year dictatorship of Hosni Mubarak, and occasionally criticized the Morsi regime that was ushered in after Mubarak was deposed in 2011 following a popular uprising, part of the Arab Spring.

The International Human Rights Network of Academies and Scholarly Societies, based in Washington DC, has actively taken up Shahin's case, and says that it will continue to defend his right to freedom of expression until all charges are dropped. Sadly, Shahin is only one of many Egyptian academics whose human rights seem to be under threat. The network has also expressed concern that, among 41,000 prisoners arrested since the coup, around 1,000 are engineers, physicians and scientists.

In April, a group of Egyptian scholars published a report on the academic victims of the unrest that followed the military coup, documenting by name and affiliation those who had been arrested or killed by the Egyptian authorities. The tally includes 1,347 student arrests and 176 student deaths. Sixteen of the deaths took place during police raids on campus. Seven faculty members have been killed, 160 placed under arrest, 20 put on parole and 25 are on the run.

What frustrated the hopes of the Egyptian Arab Spring? Morsi, an Islamist — who happens to hold a PhD in materials science — only narrowly beat his secular opponents to become the country's first democratically elected president. But moves he made to empower the Muslim Brotherhood alienated many. He proved unable to control the economy, and the country descended once again into political chaos.

Impatient for change after the revolution, in September 2011 thousands of university faculty staff from around the country demonstrated in Cairo, demanding that university leaders — all appointed directly by Mubarak — be replaced. A system allowing faculty members to elect their own rectors and deans was introduced. The protesters had also called for police to be banned from campuses unless explicitly invited by university administrations — a reaction against the oversight of campuses by state security guards during Mubarak's rule.

The new era has reversed both of these reforms. Almost immediately, police moved onto campuses to disrupt frequent demonstrations there against the regime, many organized by Islamist sympathizers angry at Morsi's removal, and many of which interrupted teaching activities. As

the new statistics show, the clashes too often ended in violence.

And last week, Morsi's successor, Abdel Fattah al-Sisi, issued a presidential decree that puts the appointment of university leaders once more under his control, a move that is widely believed will allow the regime to oust any supporters of the Muslim Brotherhood who have been elected to the posts.

Democracy cannot be built in a day. Egyptian universities have on occasion been rather naive in their response to their new freedoms. The election of academic leaders by faculty members is common in Europe, but is increasingly being phased out as its obvious weakness — that rectors might be appointed on the basis of popularity or in exchange for favours, rather than on competence — has become apparent. More usually, rectors are selected by a university board, and faculty input is indirect. Still, faculty election is better than crude political appointment.

Few campuses in the world exclude police. Greece, for example, responding to police violence against protesting students during its own previous military dictatorship, banned police entry to its universities in 1982. But as its democracy matured, its universities became more worried about the organized crime that had moved into the safe haven of the campuses — so the law was reversed in 2011. Egyptian police have been unacceptably brutal in dealing with some protesters.

Right now, academic freedom in Egypt looks to be in great jeopardy. Academics outside the country can only look on in despair and hope that the strategy of the human-rights network that represents them can at least win its battle for fair treatment of the academics detained for expressing their opinions. ■

STAP retracted

Two retractions highlight long-standing issues of trust and sloppiness that must be addressed.

This week, *Nature* publishes retractions of two high-profile papers that claimed a major advance in the field of stem cells (see page 112). Between them, the two papers seemed to demonstrate that a physical perturbation could do what had previously been achieved only by genetic manipulation: transform adult cells into pluripotent stem cells able to differentiate into almost any other cell type. The acronym STAP (stimulus-triggered acquisition of pluripotency) became instantly famous.

Soon after the papers were published on 30 January, cracks appeared

in the edifice. As various media outlets including *Nature's* independent news team reported, errors were found in the figures, parts of the methods descriptions were found to be plagiarized and early attempts to replicate the work failed.

The problems that initially emerged did not fundamentally undermine the papers' conclusions. Moreover, replication of such work is not necessarily straightforward or quick, and the ability to use some techniques can be very sensitive to aspects of the experimental protocol.

Nevertheless, the RIKEN research centre in Japan, one of the institutions in which most of the work was carried out, promptly organized an inquiry and found inadequacies in data management, record-keeping and oversight (see go.nature.com/2vrjxs). One author was found guilty of misconduct — a charge that RIKEN reaffirmed following an appeal (see go.nature.com/tnxuhy).

We at *Nature* have considered what lessons we can derive from these flaws. When figures often involve many panels, panels duplicated between figures may, in practice, be impossible for journals to police routinely without disproportionate editorial effort. By contrast, image manipulation is easier to detect. Our policies have always discouraged inappropriate manipulation. However, our approach to policing it was never to do more than to check a small proportion of accepted papers. We are now reviewing our practices to increase such checking greatly, and we will announce our policies when the review is completed.

But only since the RIKEN investigation has it become clear that data that were an essential part of the authors' claims had been misrepresented. Figures that were described as representing different cells and different embryos were in fact describing the same cells and the same embryos.

All co-authors of both papers have finally concluded that they cannot stand behind the papers, and have decided to retract them.

The papers themselves have now been clearly watermarked to highlight their retracted status, but will remain hosted on *Nature's* website, as is consistent with our retraction policy. (In our opinion, to take down retracted papers from journal websites amounts to an attempt to rewrite history, and makes life needlessly difficult for those wishing to learn from such episodes.)

We at *Nature* have examined the reports about the two papers from

our referees and our own editorial records. Before publishing, we had checked that the results had been independently replicated in the laboratories of the co-authors, and we regret that we did not capture the authors' assurances in the author-contributions statements.

We have concluded that we and the referees could not have detected the problems that fatally undermined the papers. The referees' rigorous reports quite rightly took on trust what was presented in the papers.

"We at *Nature* have considered what lessons we can derive from these flaws."

For more than two years, *Nature* has been publishing articles that highlight the problems that can arise in the analysis and publication of the results of biomedical research.

We have also launched several initiatives to improve our own rigour. For a collection of relevant content, see go.nature.com/huhbyr.

Underlying these issues, often, is sloppiness, whether in the handling of data, in their analysis, or in the inadequate keeping of laboratory notes. As a result, the conclusions of such papers can seem misleadingly robust. Another contributory factor lies in selection bias behind the data presented, whether implicit because the experiment was not randomized or blinded, or explicit in the deliberate selection of data that, usually with honest good intentions, are judged to be representative. (This is not to say that randomizing and blinding is always required, but more is needed than currently occurs.)

A manifestation of these problems has been a growth in the number of corrections reported in journals in recent years. It is hoped that the extension of our methods sections, the addition of a checklist intended to improve the standards of reporting, and our use of statistical advisers will reduce these problems in *Nature*.

In short: although editors and referees could not have detected the fatal faults in this work, the episode has further highlighted flaws in *Nature's* procedures and in the procedures of institutions that publish with us. We — research funders, research practitioners, institutions and journals — need to put quality assurance and laboratory professionalism ever higher on our agendas, to ensure that the money entrusted by governments is not squandered, and that citizens' trust in science is not betrayed. ■

Science and war

As the centenary of its outbreak approaches, Nature looks back on the First World War.

Safe in the twenty-first century, it is easy to look back at the terrible events of 1914–18 and wonder how the world turned on itself with such ferocity. Despite a century of continued conflict, the images of the First World War remain branded on our collective consciousness — the trenches, the barbed wire, the gas masks, the mud, the misery, the slaughter on an industrial scale.

The Great War was more than a clash of armies. It was a fight for supremacy in Europe and a battle to harness applications of science and technology. For the first time, machines gave the bulk of the advantage to the defenders. Science set about correcting that — an effort that climaxed in fire and fury with the dropping of atomic bombs in 1945.

Almost a century since the war broke out, *Nature* this week publishes intriguing takes on the conflict. In a Comment on page 25, Patricia Fara analyses the implications of the wartime move to recruit women into laboratories and factories. And on page 28, David Edgerton applauds writer Taylor Downing's effort to delve beneath the clichés of history and unpick how the conflict built on science from many fields. Much of that work was described in this journal, and *Nature* this month delves into its treasure trove of an archive to publish a collection

of articles from the time, including editorials, news, correspondence and book reviews, available at go.nature.com/zhlclo. Most are directly relevant to the war, but some report on other events that have entered history: the Antarctic voyage of explorer Ernest Shackleton, for instance, and work on "gravitation and the principle of relativity" presented by one "Prof. A. Einstein".

Others give a flavour of academic life. Surprisingly (or not), little has changed. There are squabbles about advertising for staff while candidates are at war; grumbles about a lack of resources (only poor-quality rubber was available for research balloons, so many burst) and a sniffy response to suggestions that scientific societies cancel their meetings. Perhaps most pertinent are articles that show how central science was to the war effort: a few days after allied troops were first gassed at Ypres, for example, a *Nature* analysis pinpointed chlorine as a probable culprit.

A warning: *Nature* at the time was rooted in the British Empire. That, and a wartime anti-German sentiment, means that some opinions and terms are not in keeping with today's enlightened internationalist attitudes. Apologies for any offence but, well, there was a war on.

The articles are bookended with striking editorials. The first, from September 1914, pointed out that Britain must restructure its industry "broadly based on science". The final piece, published days after the Armistice in 1918, presciently warned that morals must advance with

➤ NATURE.COM
To comment online,
click on Editorials at:
go.nature.com/xhunq

scientific knowledge, "for it is possible to conceive of a time when the forces at man's disposal will be so strong that a hostile army or an enemy's city may be destroyed almost at the touch of a button". The war to end all wars was only the beginning. ■



Academy 'reform' is stifling Russian science

Insufficient funding, more bureaucracy and an inefficient government funding system are sapping the life from Russian research, says Alexey Yablokov.

It is now a year since Vladimir Putin's government announced sweeping reforms of the Russian Academy of Sciences, stripping away its independence and placing it under the control of a new civil agency.

How are things going? Not well. Unfortunately, some of the gloomy predictions of critics at home and abroad that the changes would stifle research and weaken Russian science seem to be coming true.

I can speak as a member of the academy who works at one of its institutes. Formally, all of these academy institutes now belong to FASO, the government agency set up to manage us. The agency handles organization and finances, but the presidium of the academy continues to manage the scientific research. That's a very silly combination.

As part of its new role, FASO demands information from institute scientists that would be funny if it were not so tragic. We are asked to strictly plan our research. For example, how many papers will we write in a year — in two years? What kind of discoveries will we make in two years? On the basis of our promises, they then give us money.

There has, of course, been a great increase in paperwork. FASO says that it needs all this bureaucracy to guarantee our funding. The government transfers science funds to FASO, which then divides them among the institutes for salaries, expeditions, equipment, research, and so on. And just like any other bureaucratic organization, FASO wants to know what it gets for the money it gives.

Not that there is much money. There are so-called mega-grants for scientific projects, but the academy receives only 30% of the budget that the government allocates to science. The remainder goes to the high-technology business area at Skolkovo near Moscow, the Kurchatov Institute (a national research centre) and other places.

It was clear that Russian science needed reform. But the situation now is ridiculous. What the government should have done was to strengthen the way science is funded, following Western examples, such as in the United States, the United Kingdom and Germany. There, when the state wants to support science, it mostly gives the money to an independent science agency, and the agency then divides it among the researchers, taking advice from the wider scientific community. The agency staff understand how science works — they understand which teams need support, and which teams say a lot but don't do much.

In my opinion, the government wanted to liquidate the academy as a distributor of independent opinions. And of course, it wanted to get its hands on the academy's huge property portfolio.

Traditionally, academy institutes, scientific stations and labs own a lot of buildings, many in very prestigious areas, such as the centres of Moscow and St Petersburg. The

government probably wants to use these properties to make a profit. They have already taken some of the buildings from the presidium of the academy, mainly on Leninsky Prospect in Moscow, where they took two floors.

We have been told that we, the academy, have one year to get used to the new system. But FASO will really be in charge in six months. It has already announced that it will cut 6,000 administrative jobs at the academy by 2018.

Before, when I had to go abroad for scientific events, my assistant would take my passport to a specific department that took care of my visa and tickets. There were at least 50 people in that department in Moscow, and I never had any problems. Just recently, I had to go to France — but when I called that department, I was told that there were only five people left. I made the arrangements myself, which wasted

time that I could have spent doing research.

That might not sound like hardship, and it is true that some scientists are spending too much time at their holiday dachas and not in the lab — but this is because we have no money for good equipment and not enough money for field trips and expeditions.

Of course, when the Soviet Union collapsed, many scientists simply stopped working and went to their dachas to grow potatoes and carrots, to have something to eat in the winter. The situation now is not so horrible, but I know that many scientists have another job elsewhere, just to earn some money on the side because their salaries are not enough.

This is all heading towards the collapse of Russian science. Right after the reform was first announced, a huge number of young scientists and mid-career researchers with prospects and connections immediately turned to the West or the East. There are more and more of them, and they are now spending more time abroad. About three years ago, more people opted to stay at home when the salaries increased a bit. The brain drain slowed down, but about six months ago, it speeded up again.

It is not too late to recover the situation. First, the government has to give us more money. Second, this money has to be distributed under the oversight of the scientific community.

I am a biologist, and biologists know that some animals are not able to reproduce in captivity. Scientists are like that, too. We are creative people — and we need conditions in which our creativity can thrive. ■

WE ARE
CREATIVE
PEOPLE — AND WE
NEED CONDITIONS
IN WHICH OUR
CREATIVITY
CAN THRIVE.

Alexey Yablokov is a councillor at the Russian Academy of Sciences in Moscow, and former environmental adviser to the Yeltsin administration. This article is based on an interview with Katia Moskvitch.
e-mail: alexey.ablokov@gmail.com

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/jzxjkk

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PLANETARY SCIENCE

Recipe for liquid water on Mars

Liquid water could form temporarily on the Martian surface in places where salts and ice coexist.

A team led by Erik Fischer of the University of Michigan in Ann Arbor used a special chamber to simulate the environmental conditions of Mars. The researchers placed a thin layer of salt grains, similar to those found on Mars, in the chamber and exposed the salt to ice, simulating Mars's polar region. The ice melted within minutes.

The findings have implications for life on Mars, the authors say, noting that some bacteria can thrive in brines.

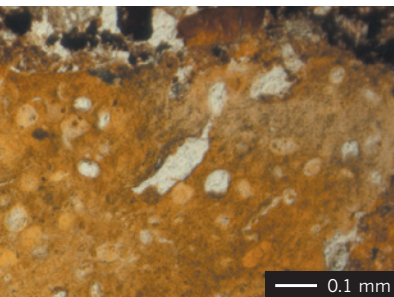
Geophys. Res. Lett. <http://doi.org/tfj> (2014)

ARCHAEOLOGY

Old faeces reveal Neanderthal diet

Traces of Neanderthal faeces — the oldest hominin waste found so far — suggest that early relatives of humans ate more plants than previously thought.

Most studies of the diet of Neanderthals, who lived 230,000–40,000 years ago, have focused on butchered bones and other signs of meat consumption, which



leave clearer evidence at archeological sites than the signs of plant consumption. Ainara Sistiaga of the University of La Laguna in Tenerife, Spain, and her team analysed sediment from the El Salt site, which dates back 50,000 years, in eastern Spain and found geochemical signals of the digested remains of both meat and plants in fossilized faeces (**pictured**).

The researchers used chemical ratios from studies of sewage pollution to argue that the faecal matter had come from humans and not from other animals.

PLoS ONE 9, e101045 (2014)

CONSERVATION

Indonesia outpaces Brazil in forest loss

Indonesia is clearing more forest (pictured) each year than Brazil, with 40% of this loss occurring in areas in which clearing is restricted or banned.

Belinda Margono at the University of Maryland in College Park and her colleagues analysed satellite images taken between 2000 and 2012 to quantify the loss of mature, natural Indonesian forest.

Natural forest cover decreased by some 60,000 square kilometres during the study period, owing to, for example, the establishment of commercial plantations and the transition of natural to managed forests.

The results raise questions about the effectiveness of conservation programmes in the tropical country, the authors say.

Nature Clim. Change <http://doi.org/tgk> (2014)



ECOLOGY

Rice pollen goes the distance

Rice pollen could be spreading further than previous studies have suggested, thanks to numerous insect species cross-pollinating the crop.

Rice is thought to be mainly self-pollinating. But Xue-xin Chen from Zhejiang University in Hangzhou, China, and his team conducted a two-year survey of rice fields across China and found that hundreds of species of insect visit rice

flowers and carry its pollen. These included the European honeybee *Apis mellifera*, which transported viable pollen grains more than 500 metres from their source. A three-year field study of genetically engineered rice showed that the bees boost gene flow, but that the proportion of unmodified seeds containing the transgene was less than 1%.

These results could have implications for controlling the spread of genes from genetically modified rice varieties.

J. Appl. Ecol. <http://doi.org/tfg> (2014)

AZWAR/EPA/CORBIS

CAROLINA MALLOL

MICROBIOLOGY

Modified microbes boost weight loss

Ingesting genetically engineered bacteria helps obese mice to lose weight by altering their metabolism.

Sean Davies at Vanderbilt University in Nashville, Tennessee, and his colleagues engineered *Escherichia coli* to produce chemicals called NAFEs, which the body converts into lipids that signal fullness. When the researchers gave these bacteria to mice on a high-fat diet, the animals ate less, gained less weight and were more sensitive to insulin than were mice that did not receive the bacteria.

These effects were maintained for at least four weeks after the bacteria were removed from the animals' drinking water, suggesting that the microbes were incorporated into the gut flora. The approach could be used as a long-term treatment for chronic conditions such as obesity, the authors say.

J. Clin. Invest. <http://doi.org/tf4> (2014)

BIOTECHNOLOGY

Halting inheritance of genetic disease

Researchers have reported a method to prevent the inheritance of diseases caused by DNA mutations in mitochondria — cellular machinery that is passed down from mother to child through the egg cell.

A possible approach to this problem is to transfer the nucleus of an egg from a woman with mitochondrial disease into a healthy donor egg, but diseased mitochondria can still carry over into the resulting embryos. Hongying Sha and Jianhong Zhu at Fudan University in Shanghai, China, and their colleagues have come up with an alternative method. They transferred the nuclei of polar bodies — small germ

cells that are created during the maturation of the egg and that have the same genome — into recipient eggs. The technique resulted in mouse pups with undetectable levels of donor mitochondria.

Polar-body transfer could one day be used to prevent mitochondrial disease, the authors suggest.

Cell 157, 1591–1604 (2014)

ASTRONOMY

Tiny asteroid in sights for capture

Astronomers have identified an asteroid small enough to be potentially suitable for NASA's mission to capture such a body.

NASA plans to use a robotic spacecraft to either net a small asteroid or grab a rock off a larger one, before dragging the body into lunar orbit for study. Michael Mommert at Northern Arizona University in Flagstaff and his colleagues studied the asteroid known as 2011 MD using the infrared camera on NASA's Spitzer Space Telescope. The team estimated the rocky body to be just 6 metres across and remarkably porous — made up of around 65% empty space, like a pile of rubble.

The mass and density of 2011 MD make it a valid candidate for a mission that grabs a whole asteroid, the third candidate identified so far.

Astrophys. J. Lett. 789, L22 (2014)

ENVIRONMENTAL SCIENCE

Less plastic in sea than expected

The amount of plastic floating in the world's oceans is between 6,000 and 31,000 tonnes — an estimate that is much less than expected given current plastic-production levels.

Andrés Cózar at the University of Cadiz in Puerto Real, Spain, and his colleagues studied more than 3,000 surface-water samples collected from oceans around the world in 2011. Nearly 90% of the samples contained plastic

SOCIAL SELECTION

Popular articles on social media

Sanger's legacy stirs up digital chatter

The late Fred Sanger, famous for work that gained him a pair of chemistry Nobel prizes, is now a social-media celebrity too — thanks to a paper by genome scientist Stanley Fields at the University of Washington, Seattle. Gaps in Sanger's publishing record had led fellow Nobel laureate Sydney Brenner to speculate earlier this year that Sanger wouldn't get research money under today's draconian funding criteria, but Fields argues that Sanger's brand of genius would still thrive. The Twittersphere wasn't wholly reassured. François Gould, a palaeontologist at Northeast Ohio Medical University, tweeted: "I appreciate this article's nuanced take on a complex issue, but [the] problem is still: most of us are not Fred Sanger."

Genetics 197, 435–439 (2014)



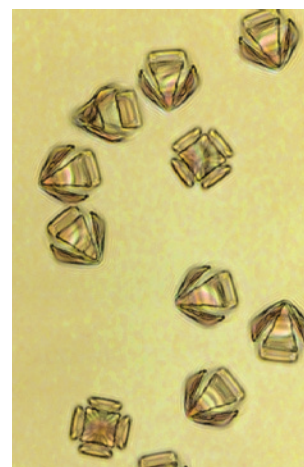
Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/hlojlr

debris, with a wide range of concentrations. The team found a lower than expected amount of plastic smaller than five millimetres in size.

The researchers think that a combination of mechanisms might explain the missing plastic: the small pieces could be sinking into the ocean interior, breaking down into even smaller, undetectable particles, or being ingested by zooplankton and small fish.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1314705111> (2014)



MATERIALS

Self-folding device grabs single cells

Tiny silicon-based grippers that can capture single cells using their self-folding arms could be useful in biological assays.

David Gracias at the Johns Hopkins University in Baltimore, Maryland, and his colleagues made their grabbing devices (pictured) out of the biocompatible materials silicon monoxide and silicon dioxide. The devices, which range in length from 10 to 70 micrometres when open,

have three or four arms that automatically fold up and around their payload. The grippers, when attached to a substrate, were able to grasp an individual mouse cell without killing it. When untethered, they could also capture red blood cells in solution.

The grippers could potentially be used *in vivo* to grab, for example, diseased cells, the authors say.

Nano. Lett. <http://doi.org/tdv> (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Presidential decree

Rectors and deans at Egypt's universities will now be appointed by the president, according to a decree issued on 24 June by President Abdel Fattah al-Sisi, who was elected in May. This reverses a law change made after the short-lived 2011 revolution that allowed faculty members to elect the heads of their institutions. New university heads are likely to be selected before the next academic year starts in September. Critics worry that the decree encroaches on universities' independence. See page 5 for more.

Forest saved

On 23 June, the United Nations Educational, Scientific and Cultural Organization (UNESCO) rejected Australia's proposal to reopen 74,000 hectares in a Tasmanian rainforest to logging. The affected region, added in 2013 to the country's 1.58-million-hectare Tasmanian Wilderness World Heritage Area, includes several Aboriginal archaeological sites. Australia's government argued that removing the protected designation from this particular part of the forest was warranted because the region is already logged and degraded, a claim that UNESCO dismissed.

Green agreements

The first United Nations Environment Assembly ended on 27 June with national representatives from around the world agreeing to push their leaders to tackle air pollution, which is linked to an estimated 7 million deaths annually. Delegates at the meeting, held in Nairobi, promised to encourage governments to set standards



NASA/JPL-CALTECH

Mars landing gear passes first test

NASA has successfully tested landing gear for use in spacecraft destined for Mars. The saucer-shaped low-density supersonic decelerator (pictured) is designed to land large loads of supplies and materials needed for long-running human missions to the red planet. It launched from the US Navy's Pacific Missile Range Facility in Kauai, Hawaii, on 28 June,

attached to a high-altitude balloon. At about 36,600 metres above the Pacific Ocean, the landing gear detached and started a powered flight, climbing 55,000 metres before starting its descent towards the ocean. It uses a new type of parachute and an inflatable Kevlar ring to slow the fall. Two more tests are scheduled for early next year.

and policies to cut emissions. Representatives also agreed on a host of other non-binding resolutions, including support for international efforts to tackle illegal trade in wildlife and improved management of chemicals and hazardous waste.

2012 and was retracted last year. But the second journal, *Environmental Sciences Europe*, conducted no further scientific peer review, its editor-in-chief Henner Hollert told *Nature*. Critics say that the study is still flawed. See go.nature.com/8x2ky9 for more.

crop yields by the amount hoped, this could feed an extra 100 million people, the study calculated — although local people will benefit only if the food is not exported elsewhere. See go.nature.com/cxuony for more.

RESEARCH

GM-study furore

A controversial retracted paper linking genetically modified (GM) maize (corn) to tumours and other diseases in rats has been republished in a different journal (G.-E. Seralini *et al. Environ. Sci. Eur.* **26**, 14; 2014). The paper was originally published in *Food and Chemical Toxicology* in

Land-sales promise

'Land-grab' deals could help to alleviate food shortages, according to a study published last week (M. C. Rulli and P. D'Odorico *Environ. Res. Lett.* **9**, 064030; 2014). Over the past 15 years, international investors have acquired some 22 million hectares of land in developing countries for agricultural purposes. If imported technologies raise

Bones reburied

The 12,600-year-old bones of a young boy that revealed the origins of the earliest Americans were reburied on 28 June in Montana, where they had been discovered. A genetic analysis of the remains, which belonged to the Clovis culture, showed that all indigenous people in the Americas had descended from a single group that migrated from Asia across a

land bridge (M. Rasmussen *et al. Nature* **506**, 225–229; 2014). The removal of the bones, which were discovered in 1968, was criticized by Native American tribes, who asked that they be returned to their original site.

AWARDS

Longitude Prize

The British public has voted antibiotic resistance as the theme of the £10-million (US\$17-million) Longitude Prize, set up by the UK government to tackle society's biggest issues. Competitors have five years to create a simple, cost-effective, accurate and rapid test for bacterial infections. That could help health-workers to avoid dispensing antibiotics when they are not needed, and so slow the rise of drug-resistant bacteria. The prize was inspired by the British government's Longitude Act of 1714, which spurred development of the first accurate marine chronometer for determining longitude at sea. See go.nature.com/wingbw for more.

BUSINESS

Exoskeleton device

US regulators have approved a device that allows people with lower-body paralysis to sit, stand and walk. The US Food



and Drug Administration announced on 26 June that it would allow marketing of ReWalk (pictured), a motorized exoskeleton made by ReWalk Robotics in Marlborough, Massachusetts. The device is the first such exoskeleton to be approved, and it is intended for use by people with some types of spinal-cord injury. It includes a brace to support the legs, motors at the joints, motion sensors and a computerized control system.

FACILITIES

Duke centre shuts

The Duke Institute for Genome Sciences & Policy in Durham, North Carolina, closed its doors on 30 June. The institute, which has obtained more than a quarter of a billion dollars in federal grants, was started by Duke University in 2002 as an

interdisciplinary centre combining fields such as bioethics and human genomics. According to the university, the institute will now be split into three units: genomic and computational biology; applied genomics and precision medicine; and science and society.

X-ray space scope

A powerful X-ray telescope will be Europe's next major space-science venture, the European Space Agency announced on 27 June. The telescope, called Athena, would be the biggest X-ray observatory ever built and is scheduled for launch in 2028. It has been a favoured project since October 2013 (see go.nature.com/r3jiil), when the agency picked the "hot and energetic Universe" as one of two themes for its next large projects, each worth €1 billion (US\$1.4 billion). Athena will image X-ray sources to study how hot gas evolves into galaxy clusters and how black holes grow.

EVENTS

Parafoil record

A test of a high-altitude balloon that is planned eventually to carry tourists has set a record for the world's highest parafoil flight. World View of Tucson, Arizona, said on 24 June that it had

COMING UP

9–11 JULY

The Infectious Diseases World Summit in Boston, Massachusetts, discusses topics such as the development of influenza vaccines. go.nature.com/spacpc

14–18 JULY

At the Eighth International Conference on Mars in Pasadena, California, subjects include Martian geology and the red planet's first billion years. go.nature.com/wloxxo

sent a scale model of its balloon system to a height of 36 kilometres on 18 June, from a launch site in Roswell, New Mexico. The parafoil deployed at 15 kilometres as the craft was descending, to help guide its landing. The 36-kilometre altitude is much lower than the 100 kilometres planned by suborbital-spaceflight tourism companies, but the ride would have the advantage of lasting several hours instead of just minutes.

El Niño brews

There is a 60% chance that an El Niño will be in effect by the end of August, and an 80% chance that it will be established by the end of December, the World Meteorological Organization said on 26 June. El Niño, which is marked by surface warming in the central and eastern equatorial Pacific Ocean, affects weather around the world, producing characteristic patterns of drought and flood. The last El Niño, in 2009–10, was generally considered moderate, but scientists say that it is too early to predict the strength of the one that seems to be brewing.

NATURE.COM

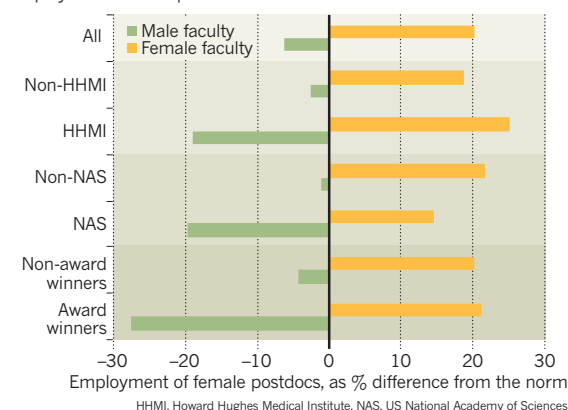
For daily news updates see:
www.nature.com/news

TREND WATCH

Male life scientists train fewer women than do their female peers, a study shows (J. M. Sheltzer and J. C. Smith *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1403334111>; 2014). The bias is most marked for lab heads who are members of the US National Academy of Sciences, are funded by the Howard Hughes Medical Institute or have won major awards. The skew could help to explain why the proportion of women drops from graduation to full professorship. See go.nature.com/qitnjp for more.

DEARTH OF WOMEN IN ELITE BIOLOGY LABS

Male principal investigators — especially the elite ones — employ fewer female postdocs than the norm.



NEWS IN FOCUS

POLICY Deep cuts hit Australia's flagship science agency **p.14**

BIOTECHNOLOGY Mexico divided over prospect of genetically modified maize **p.16**

MEDICINE Electrical jolts to organs might treat an array of ills **p.18**



STEM CELLS Concerns over California's bold approach **p.19**

JAMES GATHANY/CDC



Work with potentially lethal viruses has to be done in stringent biosafety conditions.

BIOSECURITY

Risks of flu work underrated

US funding agency disagreed with university's assessment of potential threats.

BY DECLAN BUTLER AND BRENDAN MAHER

A controversial influenza study was run in accordance with new US biosecurity rules only after the US National Institute of Allergy and Infectious Diseases (NIAID) disagreed with the university's assessments, *Nature* has learned. While some biosecurity experts express shock that the work could initially have been deemed exempt, others say that the case shows that the system's checks and balances are working.

The University of Wisconsin–Madison initially ruled that the project did not fit the criteria for dual use research of concern (DURC); which includes research that could be misused to threaten public health, national security or crops or livestock. But the NIAID disagreed.

The project, led by virologist Yoshihiro

Kawaoka, is an example of 'gain-of-function' flu research, in which influenza viruses are engineered to alter properties such as the range of hosts, transmissibility and virulence. Such research is controversial because it could be used to generate flu strains with pandemic potential. The study aimed to establish the likelihood of a 1918-like pandemic virus emerging from current avian strains, and to investigate what made the 1918 H1N1 virus — which killed some 40 million people — so virulent.

The team used genes from wild avian H1N1 viruses that coded for proteins resembling those of the 1918 virus to construct new viruses. The new viruses were not only able to spread between ferrets — the best current model for human flu transmission — they were also more virulent than the original avian viruses. The team therefore argues that there is a high risk

that a 1918-like virus could emerge naturally¹.

The NIAID has funded the six-year project since 2009. But when government introduced the new rules in March 2012, the agency asked the university to check whether the project fell under these regulations.

Research falls under the DURC rules if it meets any of seven criteria; for example, if it increases the stability or transmissibility of the agent or toxin (see go.nature.com/pwryge). But when the university reviewed the proposal in July 2012 and the progress report the following November, it concluded that the project did not meet any of the criteria. It argued that the research would not add risk beyond that done to reconstruct the 1918 flu virus in 2005 — nor would it generate more virulent viruses.

But in a letter to the university on 26 February 2013, Carole Heilman, director of the ►

► NIAID's Division of Microbiology and Infectious Diseases, says that the agency "respectfully disagrees" with these assessments. "This is the final determination," she wrote, instructing the university to develop the required risk-mitigation plan. That plan was approved in April 2013.

Nature obtained the review by the biosafety committee and the risk-benefit assessment of the university and sent them to biosafety and biosecurity experts. Richard Ebright, a molecular biologist at Rutgers, The State University of New Jersey in Piscataway, says that the research "unequivocally" met four of the seven DURC criteria. And David Relman, a member of the board that drafted the criteria and a microbiologist at Stanford University in California, says that he was surprised that the project was not initially seen as DURC, calling it a "poster child" for such research.

But Adolfo Garcia-Sastre, director of the Global Health and Emerging Pathogens Institute at Mount Sinai Hospital in New York is reassured by the process. "If anything, this means that the system is working," he says. "There are a lot of double- and triple-checks being conducted before experiments are approved, and this has been the case for these experiments."

The DURC regulations were introduced after groups led by Kawaoka and Ron Fouchier of the Erasmus Medical Center in Rotterdam, the Netherlands, managed to engineer potentially pandemic forms of the H5N1 avian flu virus^{2,3}.

Rebecca Moritz, a programme manager at the University of Wisconsin—Madison's Office of Biological Safety, says that the university had long discussed DURC issues, but had no official framework in place to assess such work, and had to put together an evaluation process at short notice after the government's rules were released.

The biosafety committee's rationale, she explains, was to compare the risks of the proposed project with those of research involving the reconstruction of the 1918 human flu virus², which had already been approved. The committee felt the new work did not pose additional risks, she says. But the committee's thinking shifted after the NIAID emphasized that the work could generate riskier versions of current bird viruses, she says.

Experts are strongly divided over the risks and benefits of gain-of-function flu research. Relman opposes such research, but says that whatever stance scientists take, everyone should agree that the process for assessing it is transparent and effective.

Oversight cannot be left to scientific agencies with an inherent interest in promoting science, he says. "What is needed is dispassionate, non-partisan rigorous oversight of this work". ■

1. Watanabe, T. *et al. Cell Host Microbe* **15**, 692–705 (2014).
2. Imai, M. *et al. Nature* **486**, 420–428 (2012).
3. Herfst, S. *et al. Science* **336**, 1534–1541 (2012).



Staff from the Commonwealth Scientific and Industrial Research Organisation facility in Hobart, Australia, demonstrated on 26 June over funding cuts.

POLICY

Australian budget hits science jobs

Research-agency staff protest over slashed spending and concerns about country's future research capability.

BY LEIGH DAYTON

Joining together in vociferous chants of defiance, nearly a thousand scientists at Australia's premier science agency protested across the country last week as it began to shed jobs in response to deep funding cuts.

The measures, announced in May's budget, include a 16% cut in funds — a reduction of Aus\$115 million (US\$108 million) — for the Commonwealth Scientific and Industrial Research Organisation (CSIRO) over the next four years. The budget, which is pending passage by the Australian Senate, also includes closures and slashed funds at other scientific organizations. The move is one in a series of steps enacted by the new government, elected last September, that have heavily targeted climate and environmental science.

The CSIRO must now decide where the budget cuts will fall. But heads of divisions — from radio science to agriculture and health — are struggling to make the required cuts while safeguarding core priorities. The first of its employees to go are 28 researchers and 3 support staff at the organization's Marine and Atmospheric Research division in Hobart and Aspendale.

The losses underscore concerns in the scientific community that the government does not support environmental and climate science. Michael Borgas, an air-quality scientist and acting secretary for the CSIRO's staff association, says that one budget measure specifically targets climate science. It includes an Aus\$20-million cut, about 69%, to the Australian Climate Change Science Program, of which the CSIRO was a major beneficiary. "Since the budget, there's been real fear for the



CSIRO STAFF ASSOCIATION

future of our world-class climate, marine and atmospheric research,” says Borgas. “Unfortunately, those fears appear to be justified.”

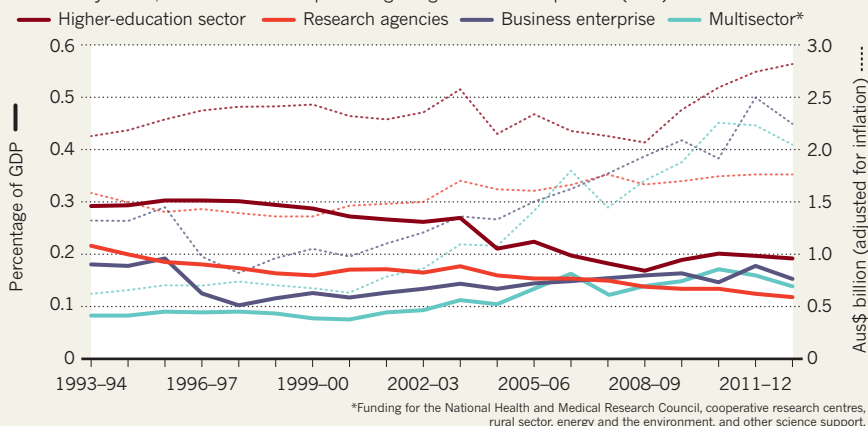
Overall, 420 of the CSIRO's 5,500 workers could lose their jobs by June 2015, according to a memo sent to staff on 14 May — the day after the budget announcement — by agency chief Megan Clark. This comes on top of 300 jobs already cut last year. By the end of the four-year process, the agency could have 1,000 fewer staff than it had in 2013, and roughly 70% of those losses would be science positions.

Research at two of the four radio observatories run by the CSIRO will be hit. Lewis Ball, head of the CSIRO Astronomy and Space Science programme, confirms that the Mopra radio telescope, 450 kilometres northwest of Sydney, will close unless alternative funding can be found. “I still do not have a confirmed budget for 2014–15, but the tea leaves aren't looking at all positive,” he says.

Southern Hemisphere tracking of NASA's interplanetary missions from the Canberra Deep Space Communication Complex, will not be affected, however. And industry

INVESTMENT DOWN UNDER

Government research and innovation spending in Australia over the past 20 years has risen in monetary terms, but has fallen as a percentage of gross domestic product (GDP).



SOURCE: AUSTRALIAN COUNCIL OF LEARNED ACADEMIES

minister Ian Macfarlane, whose portfolio includes science, confirms that funds for Australia to co-host the Square Kilometre Array, the world's largest radio telescope, are secure until 2017.

Macfarlane rejects claims that the government does not support science. He says that it is “investing strategically” in projects that will have long-term benefits for Australia. But Simon McKeon, chair of the CSIRO board, says that he is “disappointed” at the depth of the cuts and what he sees as ignorance of the long-term support that good research requires. Australia risks losing research capability, he warns.

Concerns are not just over the future of the CSIRO. Spending at four other government research agencies will also drop. One of these, the Australian Institute of Marine Science, will lose nearly Aus\$7.8 million from the Aus\$33.2 million it received last year. The Australian Research Council, which funds non-medical research, and several other research agencies will also take multimillion-dollar hits.

Environmental science is especially at risk, with about a dozen agencies set to be eliminated. But opposition in the Senate may yet save the Australian Renewable Energy Agency and the profitable Aus\$10-billion Clean Energy Finance Corporation, established in 2012 to support start-up companies. The closures are in addition to Prime Minister Tony Abbott's decision — just days after his election victory — to abolish the independent Climate Commission, a public-information body, and to draft

legislation to dismantle the Climate Change Authority, which advises the government on issues such as cuts in emissions.

The government has bought in to climate-change denial — hook, line and sinker, says immunologist Peter Doherty, a Nobel laureate and fellow of the Australian Academy of Science. Medical research, however, has not been similarly savaged: in 2015, the government aims to set up a Aus\$20-billion Medical Research Future Fund. But Doherty argues that the venture is poorly conceived because it will

“There's been real fear for the future of our world-class climate, marine and atmospheric research.”

be paid for by patient levies and funds from health-care budgets. The government “is at best naive about how science works and, at worst, indifferent to — or even contemptuous of —

science and scientists”, he says. As many researchers look for new jobs and Australia's science expenditure looks likely to drop below Aus\$8.5 billion for the first time in five years (see ‘Investment down under’), the government's chief scientist Ian Chubb warns that there is a “distinct possibility” that the nation will lose its position in international science. The problem is a lack of “a strategic vision for science”, says Chubb, who was not consulted about the budget. “Compared to the rest of the world, we're an outlier. We risk being left behind.” ■



TOP STORY



Top labs employ fewer female researchers
go.nature.com/ufi8wx

MORE NEWS

- ‘Grabbed’ land has the potential to feed millions in the developing world
go.nature.com/cxuony
- Running helps to cure ‘lazy eye’
go.nature.com/o6z3ez
- Earliest skeletal animals were reef builders
go.nature.com/tjlesm

NATURE PODCAST



Archaeopteryx's feathered trousers; teenage drinking; and how the Great War helped women into science
nature.com/nature/podcast

THOMAS BARWICK/STONE/GETTY



KARLA GACHET/PANOS

Maize is central to Mexican cuisine, culture and economy, but the formerly self-sufficient country now imports about one-third of its supply.

AGRICULTURE

GM maize splits Mexico

Legal challenge to transgenic crops has created a rift in the country's scientific community.

BY LAURA VARGAS-PARADA

The Mexican scientific community has been torn apart by a legal battle over transgenic maize (corn). Almost a year after activists challenged scientists' right to plant experimental genetically modified (GM) varieties of the crop that is a staple and symbol of Mexico, maize research is still being stymied by a legal stalemate.

On 5 July 2013, a coalition of activist groups filed a class-action lawsuit to stop the Mexican government granting permits to plant GM maize. That September, a judge ordered a halt to experimental and commercial planting until a final verdict is reached — a resolution that could take months or years.

The lawsuit and ruling have thwarted the plans of multinational companies such as Monsanto, DuPont Pioneer and Dow AgroSciences, which have lobbied for more than a decade to sell their GM maize varieties to Mexican farmers. But they have also stalled public-sector biotechnology researchers who say they are close to producing GM maize strains

tolerant to drought and frost, and other varieties with a reduced need for herbicides and fertilizers. These researchers complain that the lawsuit threatens to derail work that could boost maize yields, reduce imports and help to protect against threats such as climate change.

"We are very frustrated, and there is a general sense of despair," says Beatriz Xoconostle, a plant biotechnologist at the Center for Research and Advanced Studies (Cinvestav) in Mexico City who leads a project to develop drought-tolerant GM maize. "We have been unable to accomplish our objectives."

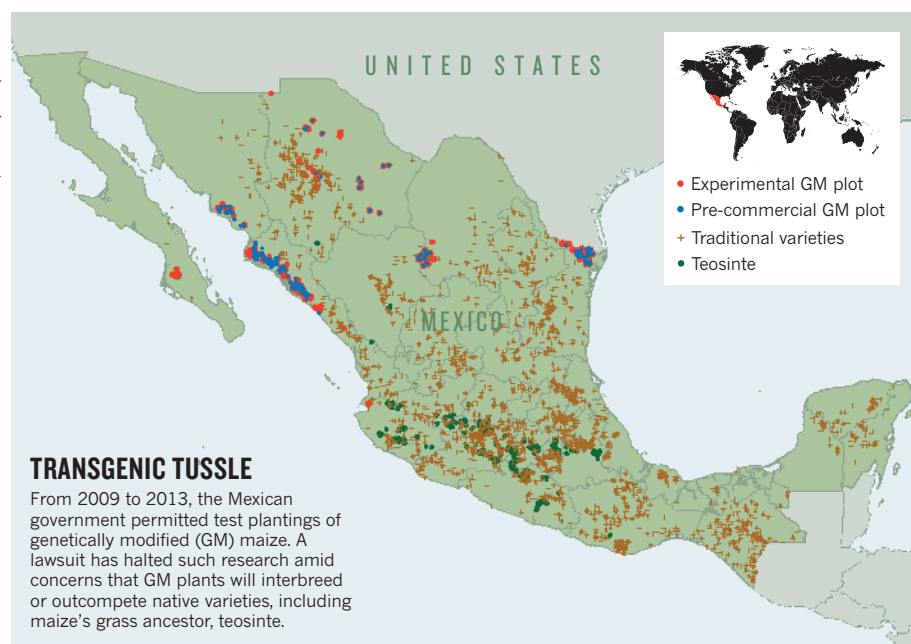
Mexico has neither embraced GM technology nor run from it. In 2010, the agriculture and environment ministries authorized commercial planting of GM cotton, and approval for transgenic soya beans followed two years later. By 2013, the country was growing 100,000 hectares of engineered crops — more than any European nation except Spain, although much less than Brazil and Argentina.

But in Mexico, the question of government approval — and public acceptance — of GM maize is much more sensitive. Fears were

stoked in 2001, after researchers at the University of California, Berkeley, reported that genetic material from GM maize had flowed into local populations of native maize cultivars (D. Quist and I. H. Chapela *Nature* **414**, 541–543; 2001). For years afterwards, even experimental plantings of GM maize were banned.

In 2009, Monsanto, Dow and DuPont Pioneer were granted approval to grow GM maize for research — as were some academic researchers (see 'Transgenic tussle'). But those programmes were halted again by the 2013 lawsuit, which was spearheaded by agronomist Adelita San Vicente of the Seeds of Life Foundation, a non-profit organization in Mexico City that opposes GM crops. The suit claimed that transgenic maize threatens the biodiversity of traditional varieties grown by subsistence farmers and smallholders throughout Mexico.

Mexico's caution over the introduction of GM maize reflects a deep desire to conserve genetic diversity in a crop that is central to the nation's identity. In the United States, the vast majority of maize is grown to feed livestock



and produce ethanol fuel. But in Mexico, 82% of white maize is grown for human consumption, often on small farms planted with traditional, rather than commercial, varieties.

"The richness of genetic diversity of maize in Mexico is invaluable," says José Sarukhán, national coordinator of the National Commission for Knowledge and Use of Biodiversity (CONABIO), a government research council created in 1992. Sarukhán and other ecologists at CONABIO worry about gene flow from GM to native varieties, and that the wide adoption of GM maize could displace them. They are also concerned that GM seed producers might take legal action against small farmers whose seed ends up containing transgenic material.

"We are not against transgenic maize, but want to raise awareness of the implications of their use, and the consequences when they mix with local varieties," says CONABIO plant geneticist Francisca Acevedo.

Once self-sufficient, Mexico now imports about a third of its maize, most of it from the United States. Scientists who oppose GM maize argue that domestic production could be boosted with irrigation and infrastructure projects, expanded agricultural-extension education programmes for farmers and careful selection of hybrid seeds and native varieties.

But government scientists, including Luis Herrera-Estrella, director of the National Laboratory of Genomics for Biodiversity at Cinvestav, see biotechnology as a crucial tool for restoring the country's maize independence. Xoconostle's project, for example, seeks to develop drought-tolerant GM maize that can also resist low temperatures. Using anti-sense RNA, she and her team have modified the plant's metabolism by inhibiting an enzyme that destroys trehalose, a sugar involved in stress response. The result, a variety called CIEA-9, requires only two-thirds of the water

needed by a normal plant. "This strategy is a way to save many of our local maize varieties," says Xoconostle.

The next stage of this research, and the last step required by Mexican law before Xoconostle can apply for a permit for commercial planting, will be to cultivate 4-hectare experimental plots of CIEA-9 to test productivity. But the team will have to wait until a final decision has been made on the lawsuit before they can go ahead.

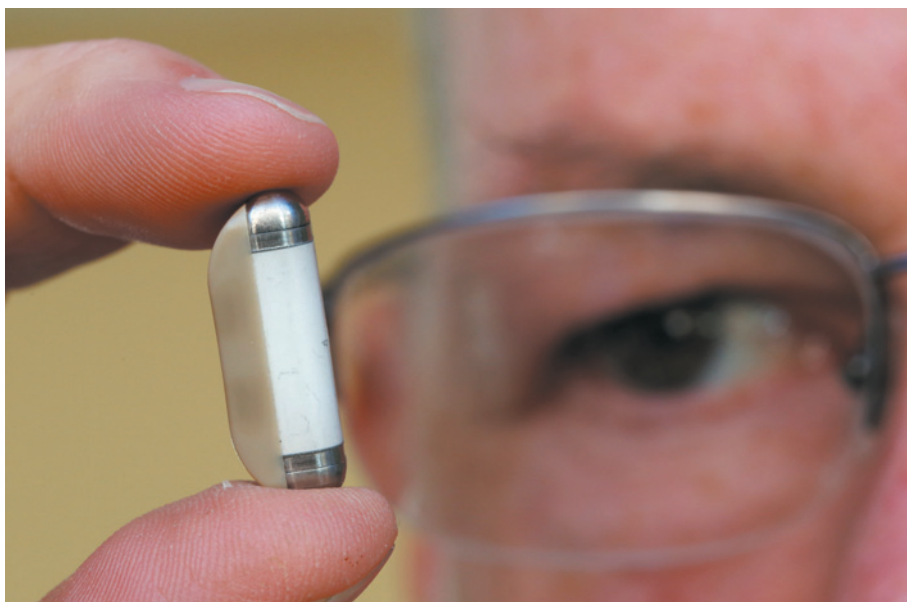
Other government crop scientists have kept their research going by moving their field trials

"We are very frustrated, and there is a general sense of despair."

to other countries. In November, Herrera-Estrella obtained a permit to plant experimental plots in Argentina. He leads a group developing transgenic maize and soya that require reduced amounts of fertilizer and compete well against weeds. The transgenic plants work by expressing a bacterial gene that codes for phosphite oxidoreductase, an enzyme that transforms the soil mineral phosphite into phosphate, which most plants need to grow and produce energy (D. L. López-Arredondo and L. Herrera-Estrella *Nature Biotechnol.* **30**, 889–893; 2012).

In Argentina, Herrera-Estrella says, regulatory requirements are "more accessible, science-based and sensible". In Mexico, more than 100 requirements must be fulfilled before a researcher can obtain a permit for experimental planting; some, he says, make the process unaffordable.

But despite her frustration with the current planting moratorium, Xoconostle does not see Mexico's stringent policies as a problem. "I am happy we have a strict law that regulates very precisely what we will allow to be grown in Mexico," she says. ■



A microregulator implant made by SetPoint Medical is designed to stimulate nerves to reduce inflammation.

TECHNOLOGY

Electroceuticals spark interest

Industry and academia invest in treating diseases by delivering electrical charges to nerves.

BY SARA REARDON

When drugs can't coax cells in the pancreas to produce insulin, or loosen arteries to reduce blood pressure, a well-placed jolt of electricity might do the trick. Spurred by decades of success with pacemakers and cochlear implants, and by advances in miniaturized technology, interest is surging in 'electroceuticals' — bioelectronic implants that stimulate nerves to treat disease.

Next week, the US National Institutes of Health (NIH) will announce a US\$248-million effort to map the body's electrical wiring and develop such devices. Pharmaceutical giant GlaxoSmithKline (GSK) has already set up a similar programme — and biotechnology companies are on the verge of bringing products to market.

On 1 May, the US Food and Drug Administration (FDA) approved a device by Inspire Medical Systems of Minneapolis, Minnesota, that stimulates airway muscles to treat sleep apnoea by regulating breathing while a person sleeps. And on 17 June, an FDA advisory committee recommended that the agency approve a weight-control device from EnteroMedics in St Paul, Minnesota. Implanted between the

oesophagus and stomach, it stimulates the vagus nerve to make a person feel full.

Scientists predict that there are many more devices to come. "The nervous system is crisscrossing our viscera to control many aspects of our organ function," says Kristoffer Famm, head of the bioelectronics unit at GSK in London. Rather than targeting cells with a drug, he says, treatments could send an electrical pulse to a major nerve to alter the commands an organ receives, and thereby control its function.

Such treatments could be more precise than pharmaceuticals, says Brian Litt, a bioengineer at the University of Pennsylvania in Philadelphia. In autoimmune diseases, for instance, it may make sense to place an electroceutical device on a well-chosen nerve rather than to blast the whole immune system with a drug. Disorders that involve targets such as the bladder, whose function is controlled by several nerve inputs, and the vagus nerve, which has a role in numerous systems including the inflammatory response and appetite, also seem ripe for electrical interventions.

"Traditional approaches of drug discovery and swallowing a pill will not be the optimal way to treat a number of diseases," says Warren Grill, a biomedical engineer at Duke

University in Durham, North Carolina, who studies electrical control of bladder function.

With that in mind, last December GSK announced a \$1-million prize for the first team to develop a miniaturized, implantable device that can read specific electrical signals and stimulate an organ to perform a specific function reliably for 60 days (K. Famm *et al. Nature* **496**, 159–161; 2013). The company has spent \$50 million on in-house electroceutical research, and it is also funding a consortium of scientists at 25 universities to develop devices that can be made available to the broader research community. Famm says that the researchers are working on electroceuticals for 20 different disorders that range from cardiovascular disease to rheumatoid arthritis and cancer. "It's a fascinating time, although we don't expect all those organs to be a slam dunk," he says.

Bioelectronic implants seem promising, but it is often unclear why they work. "Right now, a lot is based on phenomenology," says Kip Ludwig, director of neural-engineering programmes at the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. "You put an electrode in the body, you stimulate, and you get an effect."

The NIH electroceuticals project, tentatively called Stimulating Peripheral Activity to Relieve Conditions (SPARC), plans to bridge the knowledge gap by taking a step backwards and focusing on the mechanisms that underlie electrical control of organ systems. Its first grants will be awarded in early 2015. Over the next six years, the agency hopes to map the nerves and electrical activity of five yet-to-be-decided organ systems and then develop electrode devices that can attach to the nerves and maintain high-resolution recording and stimulation interfaces with them for decades without causing damage.

The most challenging task will be teasing apart the hundreds of signals that run to and from each organ. The goal is to build devices that target only the signal that elicits a desired effect, and not those that could alter functions in other parts of the body, says Litt. It is a mammoth task, he adds: "It's like putting a device across a highway, and trying to figure out, by looking at the cars passing, which will get off at which exit." ■

CORRECTION

The News Feature 'Big players' (*Nature* **510**, 204–206; 2014) erred in saying that Mario Hamuy was the only university graduate in Chile in 1982 interested in pursuing graduate studies in astronomy. He was one of just a handful of such students. The article also misnamed the organization that runs the Paranal Observatory — it should have read the European Southern Observatory's Paranal Observatory.

PATRICK T. FALLON



HOPE ON THE LINE

A decade ago, voters in California changed the biomedical research landscape by directly funding embryonic stem-cell research. Now the organization they created needs a hit to survive.

BY ERIKA CHECK HAYDEN

On a brilliant day in April, tens of thousands of baseball fans stream past Jonathan Thomas's office towards AT&T Park for the first home game of the San Francisco Giants 2014 season. Thomas's standing desk faces away from the window, but the cheering throngs are never far from his mind.

Thomas chairs the board of the California Institute for Regenerative Medicine (CIRM), the US\$3-billion agency hailed by scientists around the world for setting a benchmark for stem-cell research funding. But scientists will not be the ones who decide what becomes of CIRM when the cash runs out in 2017. Instead, it will be the orange-and-black-clad masses walking past Thomas's window. And to win their support, Thomas knows that the agency needs to prove that their collective investment has been worthwhile. "We need to drive as many projects to the patient as soon as possible," he says.

Californians voted CIRM into existence in 2004, making it the largest funder of stem-cell work in the world. The money — the proceeds of bond sales that must be repaid with \$3 billion in interest by taxpayers

— helped to bring 130 scientists to the state, and created several thousand jobs there. It has funded research that led to the publication of more than 1,700 papers, and it has contributed to five early clinical trials.

The institute has navigated a difficult path, however. CIRM had to revamp its structure and practices in response to complaints about inefficiency and potential conflicts of interest. It has also had to adapt its mission to seismic shifts in stem-cell science.

Now, ten years after taking off, the agency is fighting for its future. It has a new president, businessman Randal Mills, who replaces biologist Alan Trounson. Its backers have begun to chart a course for once again reaching out to voters, this time for \$5 billion (with another \$5 billion in interest) in 2016. And it is under intense pressure to produce results that truly matter to the public.

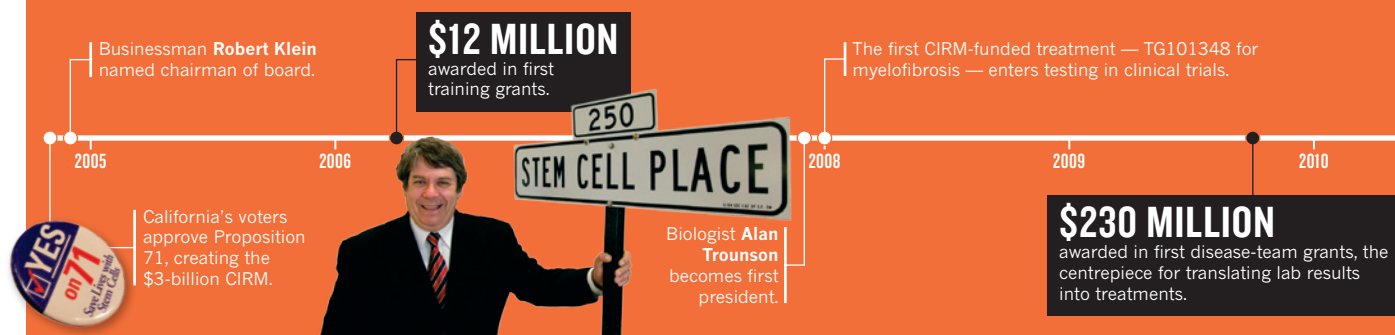
Whether or not CIRM succeeds, it will serve ►

Patient advocates and parents at a 2012 meeting in which US\$100 million in CIRM grants were approved.

► **NATURE.COM**
To see Californians speak out on CIRM's prospects, visit:
go.nature.com/t7ylv3

STATE OF FUNDING

Since its birth in 2004, the California Institute for Regenerative Medicine (CIRM) has developed a unique mechanism to support research on stem cells.



as a test bed for innovative approaches to funding. It could be a model for moving technologies to patients when conventional funding sources are not interested.

Much of what is celebrated and lamented about CIRM can be traced back to the Palo Alto real-estate developer who conceived of it: Robert Klein. Although officially retired from CIRM — he chaired the board from 2004 to 2011 (see ‘State of funding’) — Klein’s office is adorned with mementos of the agency: a commemorative shovel from the groundbreaking of a CIRM-funded stem-cell research centre, and a photo of him with former governor Arnold Schwarzenegger at the ribbon-cutting ceremony.

It was Klein’s idea to ask voters to support stem-cell research in 2004, through a ballot measure called Proposition 71. When he succeeded, CIRM instilled a kind of euphoria in stem-cell scientists, who were at the time still reeling from a 2001 decree by then-President George W. Bush that severely limited federal funding for embryonic-stem-cell research. California’s commitment removed this roadblock and revealed that many in the state and the country supported the research.

“Before CIRM, the conversation was about where to draw restrictions on embryonic stem-cell research,” says Sean Morrison, a stem-cell and cancer biologist at the University of Texas Southwestern in Dallas. “Once the California voters supported CIRM, the conversation shifted to how the rest of the country could keep up.” Indeed, five other states went on to set up stem-cell-research agencies.

PLANNING THE SEQUEL

One afternoon, Klein doodles on a legal pad as he lays out the case for ‘CIRM 2’. He writes the word ‘scientists’, then draws a box around it. Science, he says, is more under attack than ever before. President Barack Obama might have loosened the restrictions on federal funding for embryonic-stem-cell research in 2009, but some presidential hopefuls — one of whom could be elected in 2016 — support reinstating the Bush-era restrictions or even an outright ban on funding the work. Klein points out that 26 states are led by legislatures that oppose the research.

Moreover, funding for the National Institutes of Health (NIH) has taken a hit and the agency closed its Center for Regenerative Medicine in April (see *Nature* 508, 157; 2014). This is the case that Klein plans to lay before the voters. “We have to protect science’s access to the full range of cellular types now,” he says. “And in doing that, we will protect the freedom of science to ethically pursue knowledge in this country outside of religious ideology.”

His focus on a ‘full range’ of cells reflects the ways in which the stem-cell field has blossomed in the past decade, changing in ways that no one expected when Klein first proposed CIRM. And the biggest breakthroughs have come from outside California. A Japanese group, for example, discovered how to create induced pluripotent stem (iPS) cells (K. Takahashi & S. Yamanaka *Cell* 126, 663–676; 2006), which can be tailor-made from patients without the use of embryos. And scientists in Oregon managed to clone a human embryonic stem-cell line by nuclear transfer (M. Tachibana *et al.* *Cell* 153, 1228–1238; 2013).

Some treatments are progressing faster outside California, too. Japan, for instance, has raced ahead of the rest of the world in testing iPS cells as therapies for conditions such as age-related macular degeneration, which can cause blindness (see *Nature* 494, 413; 2013).

But in California, researchers are making nerve, heart, eye and skin cells from iPS cells and embryonic stem cells — a range of work rare for a single state — and they aim to test many of these in humans. They are developing drugs against cancer stem cells, which are thought to perpetuate the disease. And they are leading the world’s only two trials of treatments that combine gene editing and cell therapy to treat HIV. They are doing all this with an unmatched infrastructure, including a network of 12 new or newly renovated facilities, and a funding pipeline that acts as a beacon to young scientists.

“Almost every country would be jealous of what they’ve got in California,” says Christine Mummery of the Leiden University Medical Center in the Netherlands. Top-flight scientists such as Clive Svendsen, formerly at the University of Wisconsin–Madison, have flocked to the state, she says. But, she adds, “they haven’t cured a patient, which is the critique”.

That critique frequently bubbles to the surface. On 22 January, Thomas and CIRM’s senior vice-president of research and development, Ellen Feigal, faced uncomfortable questions from the state’s Citizens Financial Accountability and Oversight Committee at a meeting in Los Angeles. Committee member Jim Lott — a health-care executive and father of a 13-year-old daughter with a spinal-cord injury — grew frustrated as he listened to the pair run through PowerPoint slides full of numbers and abstract language touting the agency’s scientific accomplishments. None of the slides told him what CIRM had done that would get his daughter out of her wheelchair. He had asked the question before, and there was still no answer. “I’m telling you, pal,” Lott told Thomas, “I would have a hard time voting for it again.”

Lott says that he knows that CIRM is doing worthwhile science, but it is not doing a good job explaining how the science fulfils the promises of cures that were made in the Proposition 71 campaign.

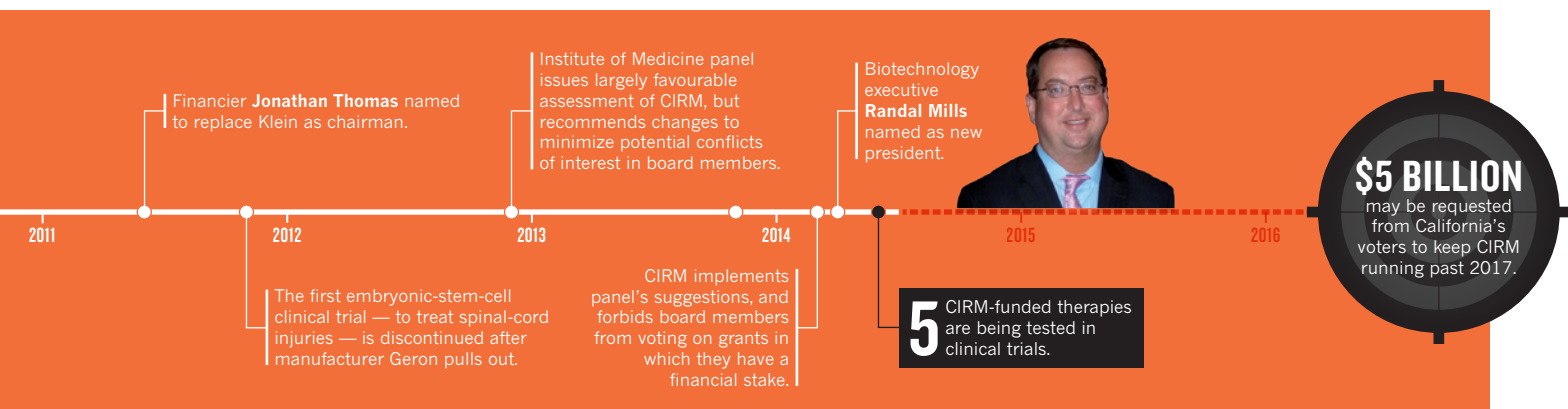
“They think that people should understand that this is good work that they’re doing, and that should be compelling enough to continue it,” he later explained. “But that’s just not a realistic view.”

And it is not just parents who have raised concerns. A 2012 report from an Institute of Medicine panel, for instance, opined that CIRM’s ambitious translational goals were “unrealistic”. Only the year before, the California company Geron of Menlo Park had halted a CIRM-funded trial that was the first major test of human embryonic stem cells after deciding to focus on cancer treatments.

Klein says that the \$34-million campaign he organized to promote Proposition 71 never promised cures in ten years. But the proposition’s supporters frequently mentioned cures — for instance, in the state’s official voter information guide, where they said that “PROPOSITION 71 IS ABOUT CURING DISEASES AND SAVING LIVES”. CIRM is now working mightily to provide evidence to sway Lott and people like him.

A big part of CIRM’s translational aspirations has been wrapped into \$450 million in disease-team awards — funds that aim to mimic the

FRAZER HARRISON/GETTY; ERIC RISBERG/AP PHOTO



CIRM work of early-stage biotechnology companies by spurring groups of scientists to move discoveries into clinical trials. To give the scientists who spearhead the teams expertise and guidance, it created a panel of industry and regulatory experts that meets with each disease team at least once a year to help to nudge them along.

And CIRM has made a series of changes to focus more on its translational mission. In May, it awarded Asterias Biotherapeutics in Menlo Park \$14 million to continue testing the product Geron had abandoned. Realizing that it would need to boost the odds of success, CIRM adjusted its portfolio to free up more cash and prioritize the projects most likely to succeed. In 2013, for instance, it cut funding for shared lab spaces and basic research, and set aside \$200 million to accelerate work by existing disease teams.

MATCHMAKER EXTRAORDINAIRE

Scientists who have benefited from these programmes say that CIRM's approach feels fresh. In 2008, for instance, it asked two groups — one led by immunologist Jeffrey Bluestone at University of California, San Francisco, the other led by the company ViaCyte in San Diego — to work together on a project to transplant insulin-producing cells derived from embryonic stem cells into people with diabetes. Bluestone worried that patients' immune systems might destroy or overreact to the cells, and had animal data showing that packaging the cells in some kind of container would prevent harmful immune reactions. So ViaCyte developed a porous membrane made from materials similar to those in medical

"I'M TELLING YOU, PAL, I WOULD HAVE A HARD TIME VOTING FOR IT AGAIN."

implants, into which it packed the cells. The company is now working with the US Food and Drug Administration to design a clinical trial.

"I don't think ViaCyte would be where it is today had there not been a partnership," Bluestone says. The NIH does not typically suggest that potential grant recipients — especially from academia and industry — pool their proposals in this way.

Irving Weissman, who heads the Institute of Stem Cell Biology and Regenerative Medicine at Stanford University in California, received a \$19-million CIRM disease-team award to work with colleagues in the United Kingdom and develop an antibody-based therapy for leukaemia. The team developed an antibody, tested it in animals, and in December received \$13 million more from CIRM for human safety trials.

Weissman says that CIRM funding has allowed science, rather than a focus on the bottom line, to drive the project. They were able to work on two potential approaches, for example, increasing the chance of a successful therapy. He says that this would be "unheard of" elsewhere. "In every biotech company, the money makes all the decisions — not the scientists."

Yet some have argued that CIRM puts too much power in the hands of scientists and others who have an inherent stake in its work. In its report, for instance, the Institute of Medicine said that CIRM's "built-in allocation of board seats to university leadership, patient advocates, and members of the biotechnology industry, for example, ensured that a high percentage of those seats would be permanently occupied by persons with almost unavoidable, conflicts of interest". In response, CIRM changed its rules so that board members whose institutions might receive money through CIRM programmes cannot vote on specific grants or funding issues.

The decision, made in March, came after years of complaints about potential conflicts on CIRM's 29-member oversight committee. This committee approves grant initiatives and awards, but one-third of it comprises administrators from academic institutes who receive the money (see *Nature* 453, 18–21; 2008). That kind of problem has plagued other efforts as well. The only other US state research funding initiative that rivals CIRM's size, the \$3-billion Cancer Prevention and Research Institute of Texas (CPRIT) has also been criticized over conflict issues, such as allegations that it awarded funds to preferred scientists without appropriate peer review (see *Nature* 486, 169–171; 2012).

Observers say that these seem to be common pitfalls of creating large state science initiatives. "Both CIRM and CPRIT have faced the reality that you can get into situations that create concerns about interest groups," says Aaron Levine, a member of the Institute of Medicine panel and a political scientist at Georgia Tech in Atlanta. On the whole, though, he adds, CIRM "has done a good job of awarding grants fairly".

Scientists who receive those grants argue that CIRM should continue to exist, but some are sanguine about the prospect of losing the funding. "In ten years more, we'll find out if CIRM's gamble pays off," Svendsen says. "But even if CIRM stops funding, we have enough preliminary data of some efficacy of stem cells that we'll continue getting funding from the NIH, wealthy donors and industry."

Thomas is trying to tap some of those sources to raise between \$1 billion and \$2 billion to help scientists to finish ongoing clinical trials if the agency closes and they run out of money. But if Klein can help it, that will not be a problem. He sees CIRM 2 as a way to tackle what he views as an ever-intensifying war on science. "If we don't take a position now," he says, "the next ten years may see a theocratic government at the state and federal level that restricts scientific research in this country for the next 50–100 years."

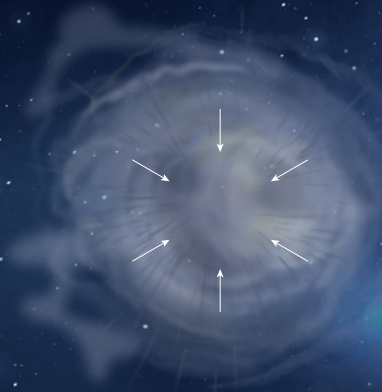
It is a shift in the message from 2004, designed to appeal to Californian pride and values. But clearly, it will resonate more if broadcast alongside ads of healthy, smiling patients who have been helped by CIRM. As one board member, former Hollywood executive Sherry Lansing, declared at a meeting in December, "We need a home run."

The point resonates back in Thomas's office as Giants fans head for the baseball stadium. They will cheer for a routine hit or a nice defensive play. But the only sure way to bring them screaming to their feet will be to slam the ball out of the park. ■

Erika Check Hayden writes for *Nature* from San Francisco.

PLANETARY STANDARD MODEL

The conventional planetary-formation theory explains how our Solar System developed more than 4.6 billion years ago.

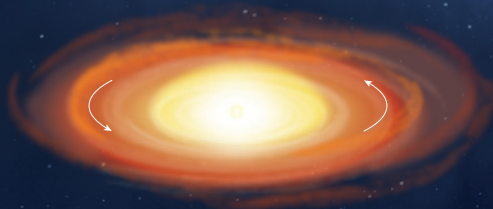


CONTRACTION

Planets are thought to form along with their stars — a process that starts with a cloud of interstellar hydrogen and helium contracting because of its own gravity.

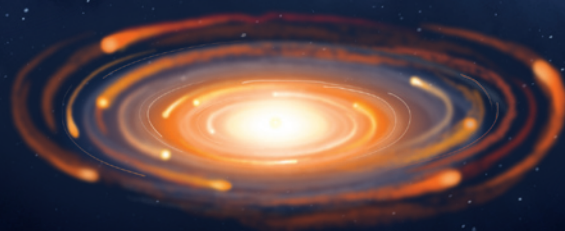
IGNITION

The cloud swirls into a flat, spinning disk with a dense blob in the centre. Temperatures and pressures at its core trigger thermonuclear fusion, and the blob begins to shine as a star.



CLUMPING

Heavy elements in the disk condense into clumps. Those in the hotter, inner part of the disk contain materials with high melting points, such as iron and rock. In the cold outer region, the clumps also include frozen 'ices' such as water, methane and ammonia.



PLANETS IN CHAOS

BY ANN FINKBEINER

The discovery of thousands of star systems wildly different from our own has demolished ideas about how planets form. Astronomers are searching for a whole new theory.

Not so long ago — as recently as the mid-1990s, in fact — there was a theory so beautiful that astronomers thought it simply had to be true.

They gave it a rather pedestrian name: the core-accretion theory. But its beauty lay in how it used just a few basic principles of physics and chemistry to account for every major feature of our Solar System. It explained why all the planets orbit the Sun in the same direction; why their orbits are almost perfectly circular and lie in or near the plane of the star's equator; why the four inner planets (Mercury, Venus, Earth and Mars) are comparatively small, dense bodies made mostly of rock and iron; and why the four outer planets (Jupiter, Saturn, Uranus and Neptune) are enormous, gaseous globes made mostly of hydrogen and helium. And because the same principles of physics and astronomy must apply throughout the Universe, it predicted that any system of 'exoplanets' around

another star would look pretty much the same.

But in the mid-1990s, astronomers actually started finding those exoplanets — and they looked nothing like those in our Solar System. Gas giants the size of Jupiter whipped around their stars in tiny orbits, where core accretion said gas giants were impossible. Other exoplanets traced out wildly elliptical orbits. Some looped around their stars' poles. Planetary systems, it seemed, could take any shape that did not violate the laws of physics.

Following the launch of NASA's planet-finding Kepler satellite in 2009, the number of possible exoplanets quickly multiplied into the thousands — enough to give astronomers their first meaningful statistics on other planetary systems, and to undermine the standard theory for good. Not only were there lots of exoplanet systems bearing no resemblance to ours, but the most commonly observed type of planet — a 'super-Earth' that falls between the

sizes of our world and Neptune, which is four times bigger — does not even exist in our Solar System. Using our planetary family as a model, says astronomer Gregory Laughlin of the University of California, Santa Cruz, "has led to no success in extrapolating what's out there".

The findings have triggered controversy and confusion, as astronomers struggle to work out what the old theory was missing. They are trying ideas, but are still far from sure how the pieces fit together. The field in its current state "doesn't make much sense", says Norm Murray of the Canadian Institute for Theoretical Astrophysics in Toronto. "It's impossible right now to account for everything," agrees Kevin Schlaufman, an astrophysicist at the Massachusetts Institute of Technology (MIT) in Cambridge. Until researchers reach a new consensus, they will not be able to understand how our own Solar System fits into the grand scheme of things, let alone predict what else might exist.

A PLANET IS BORN

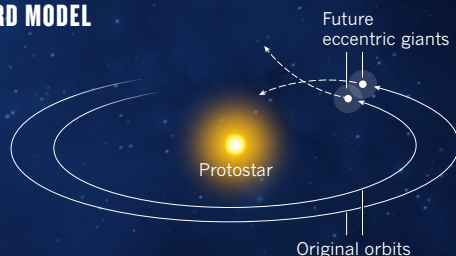
In the search for an overarching theory, astronomers do agree that core accretion has some things right: planets are leftovers from the birth of stars, a process in which interstellar

STARS: PIXELPARTICLE/SHUTTERSTOCK;
ILLUSTRATION: JASIEK KRZYSZTOFIAK/NATURE

BEYOND THE STANDARD MODEL

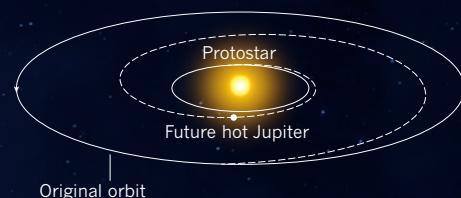
INTERACTION

If two growing planets have a close encounter, their mutual gravity could sling them off in odd directions — perhaps turning them into some of the 'eccentric' giant exoplanets seen in many systems.



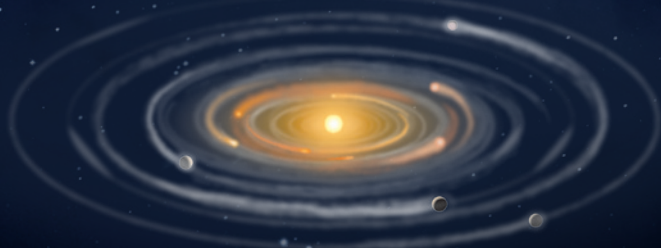
MIGRATION

A gas giant might lose orbital energy and spiral inwards because of friction with gas in the disk. If it eventually stops migrating, it will become a 'hot Jupiter' orbiting very close to the star.



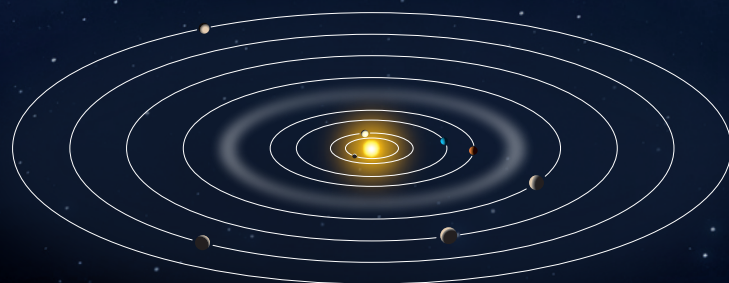
GROWTH

The solid clumps collide and join up. The bodies grow fastest in the outer part of the disk because of the ice in that region; objects there are soon so big they can also pull in gas from the disk. Friction forces the infant planets to orbit in the disk's plane, in near-perfect circles.



FINAL FORMATION

Wind from the newborn star sweeps away the disk's remaining gas. The solid clumps coalesce into several full-sized planets: rock-iron close to the star and gas giants farther out.



clouds of hydrogen and helium gas contract until their cores grow dense and hot enough to ignite (see 'Planetary standard model').

Some hydrogen and helium does not fall straight into the newborn star, but instead swirls around it, forming a thin, flat disk that orbits the star's equator. Carried along with this gas are tiny solid grains of heavier elements such as carbon, oxygen, nitrogen, silicon and iron, all made in earlier generations of stars. As the disk cools, electrostatic charges stick these grains together to form loose conglomerates that eventually grow into kilometre-scale bodies known as planetesimals. At that point gravity takes over, and the planetesimals collide, fragment, mash together and grow into full-sized planets. As that happens, friction with the surrounding gas forces them into almost circular orbits.

This core-accretion process happens throughout the disk but has different results in different locations. Towards the centre, the only grains that can survive the heat from the newborn star are materials with high melting points, such as iron and various minerals

— essentially, rock. The result is an inner system of rock-iron planets, limited to an Earth-mass or less by the disk's relative scarcity of solid materials.

Farther away from the star, however, the disk is cool enough to preserve ices that are much more abundant than iron and rock, and that accrete readily on the planetesimals. Once the planetesimals grow to maybe ten times the mass of Earth, they can start pulling in the surrounding hydrogen and helium, quickly accreting into Jupiter- and Saturn-like gas giants tens or hundreds of times Earth's mass. They stop growing only when they have cleared all the gas from their orbits.

SPACE ODDITIES

This is also where the standard theory of planetary formation stops, mainly because it fits our Solar System so well: rocky planets on the inside, gas giants on the outside. But in 1995, when observers in Switzerland reported¹ the discovery of the first unambiguous exoplanet in orbit around a Sun-like star, it was clear that the standard model had left something out. Precise measurements of the radial velocity of the star 51 Pegasi showed minuscule repeated

changes caused by a planet's gravitational pull. The data showed that the planet's mass was 150 times that of Earth, or nearly half that of Jupiter. This clearly put it in the gas-giant category. Yet the planet, 51 Pegasi b, orbited its star every four Earth days at a distance of just 7.5 million kilometres, or 0.05 astronomical units (1 AU is the distance between Earth and the Sun). This is much smaller than the 0.47-AU orbit of Mercury, and puts the planet in a region where the temperature of the gas disk during formation would have been about 2,000 kelvin, much too hot for solid ice and gases. "It was like, 'What! We weren't even looking for that,'" says Derek Richardson, an astronomer at the University of Maryland at College Park.

Astronomers called it a hot Jupiter. They soon turned up a family of such giant exoplanets between one-third and ten times the mass of Jupiter, orbiting between 0.03 AU and 3 AU from their stars. And there were other oddities: WASP-7b orbits its star's poles instead of its equator; the orbit of HD 80606b is highly elliptical, ranging from 0.03 AU at one end to 0.8 AU at the other; HAT-P-7b's orbital direction is opposite to its star's spin.

NATURE.COM
Learn more about
the Kepler planet-
finding mission:
go.nature.com/b1r8qm

By 2000, astronomers had found 30 exoplanets; by the end of 2008, 330. Then NASA launched Kepler, which spent the next four years searching for exoplanets in a single patch of sky containing some 150,000 Sun-like stars. Kepler identifies planets by detecting the slight dimming in a star's light that occurs when an object passes in front of it. This 'transit' method can find planets much smaller than the radial-velocity technique can, giving astronomers a chance to detect other Earths. Kepler has now found 974 exoplanets, with 4,254 further candidates waiting for confirmation by ground-based measurements. If all of Kepler's candidates are confirmed — and they do tend to be — then the techniques taken together will have found well over 5,000 exoplanets.

Kepler's planets run in odd systems. The Kepler-56 system, for example, has two planets, of 22 and 181 Earth masses, both orbiting at 45° to the star's plane. In the Kepler-47 system, two planets both orbit a binary star. Kepler-36's planets are closer together than any others yet seen: they orbit the star every 14 days and 16 days, respectively. One is rocky and is eight times as dense as the other, which is ice. "How did they get so close together?" wonders Richardson. "And how are they so different?" Kepler-11 is orbited by six planets, five of which are among the smallest and least massive ever found. Their densities, says David Charbonneau of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, "are shockingly low, they must be mostly ice or have significant gas envelopes" — yet all five are tucked in together within 0.25 AU of their star.

NOT LIKE THE OTHERS

Kepler's biggest surprise has come from statistical summaries of its findings. The planets seen so far can be said to fall into three categories: hot Jupiters; giant planets with idiosyncratic orbits; and super-Earths. Worlds in this third category are generally found in compact systems of two to four planets each, orbiting their stars at distances from 0.006 to 1 AU in periods ranging from more than 100 days down to hours. Although there are no super-Earths in our Solar System, they orbit at least 40% of all nearby Sun-like stars, which makes them the most common type of planet found. "The hot Jupiters are freaks, less than 1%," says Joshua Winn, a physicist who studies exoplanets at MIT. "The long-period eccentric giants are maybe 10%. The 40% — that makes you wonder."

The question is how to account for all this planetary-system diversity. In general, astronomers begin with the standard core-accretion theory then add in processes that probably did not play out in our own Solar System.

To explain hot Jupiters, for example, they suggest² that the planets did not stick around at their birth place in the cold outer reaches of stellar disks. Instead, the infant giants spiralled

inwards as viscous gas in the disk slowed their orbits. At some point, for reasons unknown, they stopped their death spirals and settled into stable orbits close to their stars. Despite the extreme temperatures, the giant planets had strong-enough gravity to keep hold of their gas.

Eccentric giants could be the result of gravitational interaction³. If several giant planets started to migrate, they might have passed one another closely enough for their gravity to sling them in crazy new directions. They could have scattered out of alignment with the rest of the system, got knocked into orbits opposite to the star's rotation or even been flung from the system entirely.

Super-Earths are harder to account for. For one thing, the term has no agreed definition, says Winn: some of the smallest, closest-in planets might actually be the stripped cores of migrating giants that came too close to their stars and got their gas blown off. "Super-Earths are probably not nice, stereotypical birds," says

"HOW DID THESE PLANETS GET SO CLOSE TOGETHER? AND HOW ARE THEY SO DIFFERENT?"

Eric Ford, an astrophysicist at the Pennsylvania State University in University Park. "Maybe some are more like penguins."

The sheer size of the super-Earth flock requires explanation. The standard theory cannot do that because in existing models, the central regions of stellar disks contain much too little material to create several close-in super-Earths. But theorists have found ways around that problem. Laughlin and Eugene Chiang, an astronomer at the University of California, Berkeley, have shown⁴ that compact systems of super-Earths can grow from disks with much greater masses, distributed closer to their stars. Murray and Brad Hansen, an astrophysicist at the University of California, Los Angeles, have also proposed⁵ a more massive disk, but one in which super-Earths are born from planetesimals that formed farther out in the disk, then migrated in before they collected into planets.

Astronomer Douglas Lin of the University of California, Santa Cruz, and his colleagues have tried to merge all the categories of planet into what Winn calls "an all-singing, all-dancing model" that can account for all the systems seen⁶. It starts by assuming that the distribution of mass in the disk will vary from system to system. After that, says Lin, it's "migration, migration, migration": all types of planet grow to full size in the middle to outer part of the disk, and then move inwards in order.

Such models are appealing, but the concept of migration, especially of the smaller planets, gives some researchers pause — if only because no one has ever seen it happening.

The necessary observations may not be possible: stars young enough to have planets migrating through protoplanetary disks are still surrounded by dust, and their light flickers, making it extremely unlikely that current methods will be able to pick out the dimming caused by a transiting planet. The theory is not settled, either. Modellers have found it hard to explain why migrating planets, big or small, would stop in the orbits that astronomers have observed. In simulations, says Winn, they don't: "the planets plop right down on the star".

Perhaps the biggest question is why our Solar System is so different. Why doesn't it contain the one kind of planet most common around other Sun-like stars? Why are there no planets inside Mercury's orbit when that's where most of the exoplanets are in other systems? Why do we have a balance of large and small planets when most other systems seem to choose one or the other but not both?

Astronomers still don't know how different we are. Observations of exoplanets are seriously biased: neither of the two main techniques would find our widely spread-out Solar System, nor are they sensitive to systems with both large and small planets. It might be that we are not unusual at all.

Future observations may give some answers. Kepler has been hobbled by a failure of the mechanisms that keep it pointing at its original target patch of sky, but last month it was approved to keep taking data. The longer it does so, the larger the exoplanet orbits it will be able to see. Ground-based programmes are starting to operate with improved instruments, some also capable of seeing planets 5 AU or more from their stars. And from 2017, NASA's planned Transiting Exoplanet Survey Satellite (TESS) will look for planetary transits across all the bright stars in the sky. The wider range of possible exoplanet candidates makes it more likely that astronomers will spot a Solar System like ours — if one exists.

Meanwhile, researchers continue to nurture their mess of models, which have grown almost as exotic and plentiful as the planets they seek to explain. And if the current theories are disjointed, ad hoc and no longer beautiful, that is often how science proceeds, notes Murray. "Life," he says, "is like that." ■

Ann Finkbeiner is a freelance writer in Baltimore, Maryland.

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. *Nature* **380**, 606–607 (1996).
3. Chatterjee, S., Ford, E. B., Matsumura, S. & Rasio, F. A. *Astrophys. J.* **686**, 580 (2008).
4. Chiang, E. & Laughlin, G. *Mon. Not. R. Astron. Soc.* **431**, 3444–3455 (2013).
5. Hansen, B. M. S. & Murray, N. *Astrophys. J.* **751**, 158 (2012).
6. Benz, W., Ida, S., Alibert, Y., Lin, D. N. C. & Mordasini, C. Preprint at <http://arxiv.org/abs/1402.7086> (2014).

COMMENT

HISTORY A refreshing study on the role of science in the First World War **p.28**



EXHIBITION Stunning show reveals art as a materials world **p.29**

BIOFUELS Policies promoting macaw palm for biokerosene production in Brazil **p.31**

SUSTAINABILITY Cookstoves: make the clean available and the available clean **p.31**

SSPL/GETTY



Women testing explosives at a factory in Gretna, UK, turned yellow from the toxic TNT and were paid one-third less than their male colleagues.

A temporary liberation

The First World War ushered women into laboratories and factories. In Britain, it may have won them the vote, argues **Patricia Fara**, but not the battle for equality.

In the early twentieth century, female scientists felt beleaguered. It is “as though my work wore petticoats”, cries Ursula, the fictionalized version of distinguished physicist Hertha Ayrton in the 1924 novel *The Call*. The real-life Ayrton was denied entry to the Royal Society in 1902 because she was married; later she struggled to make the British government’s War Office consider her design for a wooden fan to protect soldiers against gas attacks. Pre-war, alongside fellow suffragettes, Ayrton had marched behind banners embroidered with scientific figureheads including Marie Curie and Florence Nightingale, but such protests

often aroused contempt rather than support.

“I do not agree with sex being brought into science at all,” declared Ayrton. “Either a woman is a good scientist, or she is not.” Ringing words — so have we yet achieved the ideal she was fighting for a century ago?

Overt discrimination is now illegal — equality of opportunity is firmly entrenched. But all over the world, traditional attitudes linger on. Glass ceilings and leaky pipelines still present tough challenges for ambitious women in science, especially at higher levels. Exposing prejudice is the first step to eliminating it. By examining the past, we can understand how we have arrived at the

present — and how to improve the future.

In Britain, where the suffragists and violent demonstrations had failed, the First World War persuaded the government that women belonged in the polling booth as well as the parlour. “Oh! This War! How it is tearing down walls and barriers, and battering in fast shut doors,” enthused a female journalist in 1915 in the *Women’s Liberal Review*. By 1918, women had helped Britain to victory by making drugs, explosives, insecticides, alloys, electrical instruments and other essential laboratory products, and by carrying out research, running hospitals and teaching students. ▶

► Yet after the war, it was almost universally assumed that female workers should give up their jobs and slip back into their previous roles as wives and mothers. Only much later did the authorities recognize the twin follies of converting highly educated men into cannon fodder and of failing to deploy female brains effectively.

FROM EMBROIDERY TO EXPLOSIVES

As men left in their thousands for the war front, women, encouraged by the suffragist movement, seized the chance to enter fields that were previously reserved for men, including science and industry. Those with little education were trained to carry out technical but relatively routine tasks — such as inspecting bombs, testing radio valves and synthesizing chemicals. Women with scientific training stepped in to fill empty positions in universities, museums, boys' schools and government departments, or volunteered to serve abroad.

For most women involved in scientific and industrial work, the war offered a temporary reprieve from domestic servitude or leisured boredom. Wartime statistics are unreliable, but the proportion of women in employment rose from less than one-quarter to more than one-third, many of them working in munitions. Old stereotypes prevailed, so women were mainly allocated boring, repetitive tasks. They were paid at lower rates than men and regarded as intellectually and emotionally inferior. Then, as now, critics focused on women's appearances, accusing them of behaving immorally by cutting their hair or wearing uniform. Even when working with dangerous equipment, female scientists were obliged to wear long restrictive skirts rather than practical trousers.

Many women were recruited into analytical chemistry because they were deemed to be well-suited to following routine recipes — but they were paid at around two-thirds of the men's rate for hazardous work. Housed in a 16-kilometre-long complex in Gretna near the Scottish–English border, women working on TNT (trinitrotoluene) and other explosives earned the nickname 'canary girls' because their skins turned yellow from the toxic environment.

In universities around the country, isolated handfuls of female researchers diverted their attention to secret military and medical projects. Highly qualified, these women had fought hard to pursue an academic career, often facing hostility from their own families and from the men who taught them. Banned from most scientific societies, they found it hard to get teaching posts. "Women high up in scientific positions, women with international reputations," protested the palaeobotanist Marie Stopes, "are shut out from the concourse of their intellectual fellows." One of the first female lecturers in science, Stopes studied coal power for the government during

the war, only later dedicating herself to sexual education and family planning.

Scientific women were keen to volunteer their expertise. "I can put all my time and energy at your service for the next 6 weeks," wrote Margaret Turner, a pharmacologist at the University College of Wales, to the government, "and am anxious to know whether the few helpers down here could not be allowed to contribute further to the needs of the country?" Little trace is left of women's

"The proportion of women in employment rose from less than one-quarter to more than one-third."

wartime activities, and details can be hard to glean. For example, in 1914, the chemist Frances Micklethwait joined a team in London making explosives and was awarded an MBE (Member of the British Empire), one of Britain's highest distinctions — but because the work was top secret, little further information has survived.

Unprecedented openings arose for a few women. Like other female curators, the palaeontologist Dorothea Bate was ineligible for an official position at the Natural History Museum in London, but a few months after the hostilities began she stepped in to replace men who had volunteered to fight abroad. As more men disappeared to the front, some never to return, Bate's role expanded, but she was still paid piecemeal, so her weekly earnings remained lower than those of less-skilled male assistants. Even after the war, pressure remained high because of the backlog of uncatalogued specimens, but it was not until 1928 that the museum permitted women to apply for jobs — the ones at salaries too low to attract men. Thirty-seven years later, Bate was still there and still on a temporary contract.

Protectionism could bring temporary

advantages. Up to 1914, the single-sex Balfour Biological Laboratory at the University of Cambridge offered a secluded environment in which women conducted high-level research in a supportive community. Crucial for the first generations of female scientists, the laboratory closed when the departure of male volunteers to the war increased the available places across the university.

Having successfully demonstrated their scientific prowess, women were no longer automatically excluded from laboratories and lecture theatres. Then and later, in emerging fields such as genetics and X-ray crystallography, a few men — including William Bateson and William Lawrence Bragg — welcomed women to their teams. Perhaps they recognized that those who had reached such an advanced level must be of an exceptionally high calibre and could be employed more cheaply than men to carry out the same work.

ENTRENCHED ATTITUDES

Like their male colleagues, the few female researchers spent the war years divorced from their own interests. In the artificial trenches dug in the gardens of Imperial College London, the pharmaceutical chemist Martha Whiteley experimented with her team of seven female assistants on explosives and poisonous gases. Almost 40 years later, in a 1953 lecture intended to inspire young women, she described examining the first sample of mustard gas, a blister agent, to arrive in Britain: "I naturally tested this property by applying a tiny smear to my arm and for nearly three months suffered great discomfort from the widespread open wound it caused in the bend of the elbow, and of which I still carry the scar." Unusually, Whiteley resumed her own research on organic analysis after the war.

Especially at the beginning of the war, the official policy was clear: women were



Physicist Hertha Ayrton (left) designed a fan to protect soldiers from poisonous gases; chemist Martha Whiteley (right) tested such gases in artificial trenches at Imperial College London.



AYRTON: SCIENCE PHOTO LIBRARY; WHITELEY: IMPERIAL COLLEGE ARCHIVES



© IMPERIAL WAR MUSEUM

Botanist Helen Gwynne-Vaughan, the first overseas commander of the Women's Army Auxiliary Corps and a commandant of the Women's Royal Air Force, painted by William Orpen (1918).

unsuitable for action overseas; their role was to keep the home fires burning. Patriotic scientists who relished the excitement and independence of travelling abroad had two main options — to volunteer for medical teams or to join the army.

Within a fortnight of war being declared, suffragist organizations had raised funds for a professionally staffed and fully equipped research hospital to be sent abroad. The offer was rejected by the War Office, which replied curtly that serving soldiers “did not want to be troubled with hysterical women”. By contrast, Britain’s allies — including Belgium, Russia and France — eagerly accepted, and British female teams worked in Serbia and Thessaloniki, Greece, throughout the war, later joined by US units. In well-kitted laboratories they carried out research into tropical diseases such as malaria and dysentery. And like their foreign colleagues, including Austrian physicist Lise Meitner and Marie Curie, British female scientists volunteered as radiologists.

These volunteers endured atrocious conditions, building hospitals from scratch, often in freezing weather, and dealing with horrific injuries and illnesses. And they continued to experience substantial discrimination. The War Office decreed that the former research botanist Edith Stoney was unfit to become head radiologist in Thessaloniki because she was a woman. They would have changed their mind, commented the doctor Isabel Emslie Hutton, had they seen

her “carry heavy loads of equipment, repair electric wires sitting astride ridge tents in a howling gale, and work tirelessly on an almost starvation diet”.

Some scientists found themselves unexpectedly recruited for military service. Helen Gwynne-Vaughan was head of the botany department at Birkbeck College, London, when she was plucked from what she described as “perhaps the only sphere in which at that time young men and women worked freely together — the laboratories of a modern university”. As the first overseas commander of the newly formed Women’s Army Auxiliary Corps (WAAC), Gwynne-Vaughan crossed the English Channel in 1917 “into a new and different world”. She immediately set about changing it, insisting that women be treated exactly the same as men, and persuading male officers that jokes such as “Would you rather have a slap in the eye or a WAAC on the knee?” would not be tolerated. After subsequently running the Women’s Royal Air Force for a year, Gwynne-Vaughan returned to her former position laden with national honours, but keen to resume her research into the genetics of fungi — an important topic at a time when refrigerators were scarce and food transport was slow.

LABORATORY POLITICS

After the armistice, women over the age of 30 quickly gained the right to vote. Nonetheless, most were deemed extravagant

opportunists who were taking the rightful wages of fathers and husbands.

As soon as the men returned, university women were forced to relinquish their wartime positions, especially in lecturing, which was seen as a male preserve. A sympathetic chemistry professor protesting to the War Committee on behalf of his female colleagues wrote: “these women gave their services, services which have not received any public recognition”. At University College London, the female academics eventually secured a tea room (a former chemistry laboratory, with a gas burner in the fume cupboard) — one brave individual ventured into the men’s common room and there was “such a sanacker-towzer of a row” that she backed out, defeated.

Although women did benefit from the post-war expansion of education and research, they still faced enormous obstacles in pursuing scientific careers. Female students, who had made up the majority during the war, were now outnumbered by men and were once again unwelcome in lecture rooms and laboratory classes. As the feminist journalist Cicely Hamilton lamented in 1935 that “we are retreating where once we advanced; in the eyes of certain modern statesmen women are not personalities — they are reproductive faculty personified.” At Cambridge, women could not formally graduate until 1948. Although some scientific societies accepted female members, separate common rooms survived until after the Second World War.

Scientific opportunities in industry opened up after the war, but married women with the vote often supported the conventional view that women belonged at home. As unemployment soared, they preferred available jobs to go to their menfolk, rather than to younger single women. Younger women were forced to take unskilled jobs at lower pay and found themselves trapped near the foot of the career ladder. As Kathleen Culhame, a disillusioned drugs researcher, explained: a male chemistry graduate could expect to progress well, but the “girl who worked side by side with him at the university is hard up and constantly humiliated ... She will be happier if she is not too enterprising because then her sense of frustration will be less.”

So, yes, the war enabled more women to enter science. But a century on we are still rooting out the discrimination that was built right into the heart of the system. ■

Patricia Fara is the senior tutor of Clare College at the University of Cambridge, UK. Her books include *Science: A Four Thousand Year History* and *Scientists Anonymous: Great Stories of Women in Science*.
e-mail: pf10006@cam.ac.uk



Soldiers load a Stokes mortar, invented in 1915 by engineer Wilfred Stokes for trench warfare.

MILITARY TECHNOLOGY

Science in the trenches

David Edgerton applauds a study of a scientific elite whose impact spanned two world wars.

Many clichés bedevil our understanding of the role of science in the two world wars. One that has lingered is the idea that the First World War was a chemists' war (poison gas, explosives) and the Second World War belonged to the physicists (radar, the atomic bomb). To judge from past pages of *Nature*, the term 'chemists' war' was not coined until 1920, and then only in reference to a future war, whereas 'physicists' war' was used to

describe mechanization, radio and aeroplanes in the early phases of the Second World War. In any case, both clichés are absurd: experts of all kinds were involved with the military in peacetime, and more so in war.

One of the great virtues of Taylor Downing's lucid and entertaining *Secret Warriors*, on experts in the British First World War effort, is that it is unaffected by such hackneyed ideas. Downing deals with aeroplanes, plastic surgery, new weapons, aerial photography and

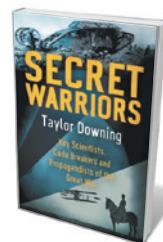
poison gas. He blithely ignores chemist and writer C. P. Snow's notorious and fatuous 'two cultures' thesis, because his book also covers the work of historians, journalists, linguists and physicians at the time.

Secret Warriors is full of interesting characters, nicely illustrating the deep interconnections between British experts and war. Take Alfred Ewing, former professor of mechanism and applied mechanics at the University of Cambridge, UK. He went on to become Britain's director of naval education in 1903, and had a key role in wartime code-breaking for the Admiralty in London. There he created what was known as Room 40, into which went code-breakers Dilwyn 'Dilly' Knox, Nigel de Grey and Alexander Denniston; they continued their work through the inter-war years and became key members of the UK government's code and cipher-cracking operation during the Second World War at Bletchley Park in Buckinghamshire.

In aviation, the connections and continuities are just as striking. The Cambridge physicist John William Strutt, Baron Rayleigh, was involved in promoting large-scale military aeronautics research from well before the First World War; Britain sent a small air force to France soon after war was declared. Aeronautical problems were rife, and solving them required many scientists, some of whom would become important advisers during the Second World War. The chemist Henry Tizard, for instance, worked on bombsights and went on to become the Royal Air Force's main scientific adviser into 1943, while the physicist Frederick Lindemann became Winston Churchill's personal adviser and a member of the wartime cabinet.

In land battles, the first tank or 'landship', an armoured vehicle with huge caterpillar tracks that took it over trenches, was the work of men associated with the military from before the war. These included the gloriously named naval architect and engineer Eustace Tennyson d'Eyncourt, the director of construction for the Royal Navy, who was a relation of the Victorian Poet Laureate Alfred Lord Tennyson.

There were also innovations and important inputs from those previously unconnected to the military, such as industrial engineers William Mills and Wilfred Stokes, inventors of the Mills grenade and Stokes mortar, respectively. Journalists and historians were brought in to conduct covert and overt propaganda — from great press barons such as Viscount Northcliffe to John Buchan, author of



Secret Warriors: Key Scientists, Code Breakers and Propagandists of the Great War
TAYLOR DOWNING
Little, Brown: 2014.

FOTOTECA GILARDI/GETTY

the classic thriller *The Thirty-Nine Steps*.

Along with that huge range of contributions to the war effort, what becomes evident in *Secret Warriors* is the easy relations between different elements of the elite to which nearly all these scientists and scholars belonged. It is fascinating to see how similar the situation was in both world wars, and how many names appear in accounts of both. At least 13 of the men mentioned by Downing were as familiar in the Second World War as in the first.

The straightforward story Downing tells is a refreshing change from older treatments of science and war, including academic ones. These derived from stories scientists wanted to have told — accounts of the military's indifference to science or biographies of conscience-stricken nuclear physicists. Such arguments have led to a lack of frankness about the actual relationship between science and war; even some literary scholars with an interest in science argue that it was only in the Second World War that science and the military became allied. Historians of science with a little more knowledge have even claimed that there were negligible links between the two before 1914. Downing has done a great service by ignoring these risible theses. The connections are long-standing, deep and important to both, and have been celebrated for centuries.

Downing does well to ignore previous accounts that stress an alleged hostility to new techniques among the British elite — a view that has severely distorted our understanding of the history of British science. But a little declinist mud has stuck to his boots, even when he is contradicted by the evidence that he presents. Thus he laments that the British military academy Sandhurst did not teach science, which might seem damning; but he notes elsewhere in the book that Woolwich, which trained engineers and artillery personnel, did. And although it might seem true that British universities were dominated by pure science and the arts, that is to ignore key points in other parts of the book, such as the importance of Cambridge engineering, or the point that many British universities were, to a significant degree, medical schools. In Britain, as elsewhere in both peace and war, science was and remains deeply connected to the military. What now needs to be explained is why so many intellectuals, especially in the sciences, sought to deny or ignore this. ■

David Edgerton is *Hans Rausing professor of the history of science and technology and of modern British history at King's College London, UK. He is author of England and the Aeroplane. e-mail: david.edgerton@kcl.ac.uk*



Vermilion, an ancient pigment, features in Edgar Degas' innovative 1896 *Combing the Hair*.

SCIENCE OF COLOUR

Hue and eye

Barbara Kiser revels in an immersive show revealing the scientific base coat to 700 years of European art.

For millennia, artists chased the blues — from Egyptian blue frit to ultramarine, azurite, cobalt blue and indigo. Derived from earthbound minerals, metals and plants, these pigments signalled the celestial, tinting the robes of painted Renaissance Virgins by the score, and the saturated sky behind the meeting of god and mortal in Titian's sixteenth-century masterwork *Bacchus and Ariadne*. The hunt for blue — and all the other hues — is a tale as bound up in science and technology as in aesthetics and tradition.

That long, complex quest is chronicled in *Making Colour* at London's National Gallery. The show peers into the luminous slicks of paint in a Monet or a Memling to reveal the chemical and physical traits of the pigments and their origins. In rooms devoted to blue, green, yellow, red, purple, and gold and silver, exemplary paintings are set next to raw materials and the pigments derived from them. Alongside are manufactured objects glazed or dyed with the same colourants, pointing to the fact that up to the nineteenth century, many pigments were by-products of dye manufacture, metallurgy and other industries.

"We are trying to tell the history of materials in European painting in very visual form,"

Making Colour
The National Gallery,
London.
Until 7 September
2014.

explains exhibition co-curator Ashok Roy. He heads the gallery's science department, which for decades

has analysed the collection — spanning key schools of European art from the fourteenth to the nineteenth centuries. Over much of that time, artists "needed spectrally powerful colours for their paintings, particularly for the often brilliantly coloured robes of religious figures", notes Roy.

A vast sunburst of colour presides at the start of the show — the 72-colour 'chromatic diagram' devised in 1839 by chemist Michel Eugène Chevreul as a way of thinking about complementary colours. Nearby is Vincent Van Gogh's *Two Crabs* (1889), an electric contrast of hot reds and acid greens that testifies to Chevreul's influence. A cluster of other colour wheels includes that of Johann Wolfgang von Goethe. The German polymath's ideas on colour were muddled, but they opened the way to Chevreul's and influenced J. M. W. Turner, whose palette, smeared with a muddy impasto of russets and greens, is showcased across the room.

Beyond, the near-black walls give a ▶

► chromatic intensity to the paintings and objects. Another kind of intensity lies in what it took to acquire, transport and process some pigments. The 'blue room' features a vast photo of an Afghan miner lugging a toddler-sized chunk of lapis lazuli — the source of the mineral lazurite, from which natural ultramarine was derived. From around 1200 to the nineteenth century, European painters imported lapis lazuli from the remote Sar-e-Sang mines in what is now Afghanistan, by way of Baghdad, Damascus, Cyprus and Italian ports. The extraction of the pigment, a chore involving lye, wax and plenty of elbow grease, failed to deter artists eager for strong blues and the cachet of costliness.

Thomas Gainsborough's 1785 portrait of theatre diva Sarah Siddons in a blue-striped dress marks how chemical experimentation provided an alternative: Prussian blue. Around 1704, the Berlin alchemist and colour-maker Diesbach was attempting to make red 'lake' pigment using potash, iron sulphate and a decoction of dried cochineal beetles. The potash, however, was contaminated with 'animal oil' derived from blood, and the result was a deep blue — one of the first modern synthesized pigments.

Pigments that were spin-offs from industry are showcased by a traverse through yellow. The mustard tint on the robe of a majolica (glazed terracotta) angel by the Renaissance Della Robbia workshop is

Majolica (glazed terracotta) angel by the Renaissance Della Robbia family.



A family of yellow pigments highlight a gown in Anthony Van Dyck's portrait, and an angel's robe (below).

based on a lead and antimony colourant — a family of yellows that sometimes also incorporates tin. These stable oxides have been used in glass-making and ceramic glazes for centuries, and they appear in works by seventeenth-century artists such as Nicolas Poussin and Anthony Van Dyck. For the decidedly worldly figure of the viscountess in *Lady Elizabeth Thimbelby and Dorothy, Viscountess Andover* (1637), Van Dyck picked out the peaks in her sumptuous gold-satin dress with lead-tin yellow; Pietro da Cortona's *Saint Cecilia* nearby features a yellow based on lead, tin and antimony.

The nineteenth century was, in Roy's words, "an explosion of colour chemistry" that crowded palettes with cobalt blue, cadmium colours and synthetic ultramarine. The sharp dabs of fresh green and olive in Paul Cézanne's 1890 *Hillside in Provence* testify to the strength, stability and brilliance of the new chromium-based viridian and the copper-arsenic compound emerald green. Premixed in tubes, these pigments helped to enable the novel practice of painting *en plein air*. Yet new did not negate old. In his 1896 *Combining the Hair* — a conflagration of pinks, brick reds and terracottas centring on the taut, flame-like hair of the seated girl — Edgar Degas deploys red earth, red lead and the mercury sulphide

vermilion. This last has been used in painting, first as the natural mineral cinnabar and then in a synthetic form, for millennia.

The chromatic trail of primary and secondary colours comes to a halt in a welter of metallics. Giovanni Girolamo Savoldo's sixteenth-century *Mary Magdalene* almost steals the show, even though the silvery hue of her cloak is a beautiful illusion crafted in lead white and lamp black. That every hue you have seen here is a dance between wavelengths of electromagnetic radiation, the optic nerve and the brain is left for last (and for more on this and many topics featured in the show, see Philip Ball's excellent *Bright Earth*; University of Chicago Press, 2001). On the way out of the exhibition, a well-crafted film on human perception awaits, including a crowd-sourced experiment. Created by neuroscientist Anya Hurlbert and backed by the London-based Wellcome Trust, this examines colour constancy (how we perceive colours as the same in different lights) by inviting visitors to assess two reproductions of a Cézanne still life under varying illumination.

But at base, *Making Colour* ushers us into a materials world. Beyond the artists' transformation of coloured earth or ground resin into luminous windows on lost worlds is a broader transformation in chemistry, exploration and commerce — a necessary underlay to the aesthetic urge. ■

Barbara Kiser is Nature's Books and Arts editor.

LADY ELIZABETH THIMBELBY AND DOROTHY, VISCONTRESS ANDOVER, 1637/© THE NATIONAL GALLERY, LONDON

KNEELING ANGEL HOLDING A CANDLESTICK, 1500-50/LENT BY THE SYNDICS OF THE FITZWILLIAM MUSEUM/© THE FITZWILLIAM MUSEUM, CAMBRIDGE

Correspondence

Brazil promotes aviation biofuels

The drive to produce biofuels that are suitable for aviation is starting to look promising (see go.nature.com/ujv5ne). For instance, the state government of Minas Gerais in Brazil has put in place policies to stimulate the cultivation, extraction and processing of the native macaw palm (*Acrocomia aculeata*), a potentially sustainable source of oil for producing biokerosene.

This tree can produce up to 6,200 kilograms of oil per hectare (see T. P. Pires *et al. Ind. Crops Prod.* **44**, 200–210; 2013) and can sequester as much as 4–5 times more carbon than other afforestation species such as *Eucalyptus camaldulensis* (see H. Suganuma *et al. J. Arid Land Stud.* **22**, 69–72; 2012).

To safeguard forests, macaw palms will be planted only on existing pastureland (estimated at 170 million hectares in 2010). Oil production from macaw palms, which could exceed the size of today's global palm-oil market, does not need to compromise food production or natural ecosystems (see go.nature.com/67tfia). The trees can be cultivated next to agricultural crops and their extensive canopy cover provides fauna with food and shelter, helping to restore biodiversity that has been lost by forest clearance.

Éder Cristian Malta de Lanes, Paulo Mafrá de Almeida Costa, Sérgio Yoshimitsu Motoike
Federal University of Viçosa, Minas Gerais, Brazil.
edercm@gmail.com

Reassess Chinese research awards

China's research-award system needs reform if it is to encourage future advances in science and technology. We believe that a move away from current government schemes — which mainly recognize large

technological projects — towards non-governmental awards for individual researchers would boost academic competition and be more effective in the long term.

Non-governmental award schemes in China do not carry the same prestige as they do in Western countries. The Chinese government gives national prizes every year to about 280 research projects and six individual researchers, on average, and regional prizes to a further 3,000 projects. Scientific societies, which are mostly affiliated with the government and led by retired government officials, offer annual prizes for some 10,000 projects. Most of these awards are intended to motivate technological invention and application. Less than 5% of prizes are bestowed for research in the natural sciences.

In our view, the Chinese government should deregulate awards related to technological inventions and applications and allow the prizes to be determined by the market. China's scientific societies also need reform to give them more authority and greater autonomy, so that their research awards can focus on academic achievements and be more influential.

Yuan-Feng Wang, Hui-Bing Xie
Beijing Jiaotong University, China.
cyfwang@bjtu.edu.cn

Cookstoves: clean up fuel on two fronts

Contrary to the impression you convey (see *Nature* **509**, 533; 2014) biomass-burning cooking technology is advancing steadily. Stoves are now more efficient and emit much less smoke, and will remain popular as long as users can access biomass, such as wood and dung, at zero direct cost. Efforts must therefore continue to make clean fuels available and available fuels clean.

In trials of new types of biomass-burning stove

— including the trial covered by the 2012 report you mention — one factor contributing to the apparent negative outcome is promoters' use of the term 'improved' to market new stoves, often without justification. This has led to the conclusion that the 'improvements' have not worked. Genuine improvements can stem only from systematic testing and assessment. Moreover, randomized controlled trials of health interventions need to follow strict criteria (see, for example, K. R. Smith *et al. Lancet* **378**, 1717–1726; 2011).

This autumn, the Global Alliance for Clean Cookstoves in Washington DC and ISO, the International Organization for Standardization, are due to finalize the first health-based emissions standards for biomass stoves. These are informed by the World Health Organization's upcoming Indoor Air Quality Guidelines. No longer will funders, non-governmental organizations, the private sector, the media and researchers have to rely on vague and unsubstantiated descriptors to judge stove performance.

Kirk R. Smith
University of California, Berkeley, USA.
krksmith@berkeley.edu

Cookstoves: a design popular in Peru

Conversion to clean-energy fuel is a priority for cooking among Peru's rural communities. This is for two reasons: because of the adverse health effects of indoor smoke from biomass-burning stoves (see *Nature* **509**, 548–551; 2014) and because using tropical hardwoods as fuel is accelerating deforestation in the Amazon.

The only viable source of clean fuel for rural communities in the Peruvian Amazon is natural gas, which has to be extracted and transported across the Andes, imposing prohibitively high running costs.

A pyrolysis cooking stove offers another solution. Its

combustion design (see go.nature.com/e7j8q1) makes it superior to the biomass-burning stoves you describe. For example, it cooks faster than an open fire but consumes half the amount of wood, generates half the smoke and can run on low-grade fuel such as small twigs and leaf litter.

At US\$180 per unit, it is costly compared with the \$15 stoves that you discuss, but the price incorporates training, monitoring and education to ensure that the stove is properly used. It is also a simple matter to convert these stoves to burn natural gas, should that become more economically viable.

Kaysara Khatun
Autonomous University of Barcelona (UAB), Bellaterra, Spain.

Brenton Ladd
Scientific University of the South, Lima, Peru.
brenton.ladd@gmail.com

Longitude Prize now an objective decision

Martin Rees writes that John Harrison “came closest to receiving the reward money” offered in the 1714 Longitude Act (*Nature* **509**, 401; 2014). This description conceals a long and painful battle between astronomers and those who argued that Harrison should be awarded the prize for his marine chronometer, designed to determine the longitude of a ship's position by using existing tables.

The astronomers of the day claimed that watching Jupiter's moons, as suggested by Galileo, was the scientific solution. But in practice, this ‘astronomic’ solution would not work under a cloudy sky.

Today, by contrast, the winner of the 2014 Longitude Prize, who we now know will be working on antibiotic resistance, will be “decided objectively — as in athletics”, as Rees points out.

Carl G. Ribbing
Uppsala University, Sweden.
cg.ribbing@angstrom.uu.se

Group size and cultural complexity

ARISING FROM M. Derex, M.-P. Beugin, B. Godelle & M. Raymond *Nature* **503**, 389–391 (2013)

A decade ago, Henrich¹ proposed group size as a driver of cultural complexity. Derex *et al.*² now present experimental results they say support this ‘group size hypothesis’ by seemingly showing that larger groups perform better than smaller groups under imitation-based cultural evolution. Our reanalysis of their experimental data, however, shows that larger groups actually perform worse than smaller groups. Thus, contrary to their claim, their data are consistent with empirical evidence discounting the group size hypothesis for non-food producing societies^{3–10}. There is a Reply to this Brief Communication Arising by Derex, M. *et al.* *Nature* **511**, <http://dx.doi.org/10.1038/nature13412> (2014).

In their argument, Derex *et al.*² assume that showing that larger groups produce complex artefacts more often than do smaller groups (see their Fig. 1) suffices for supporting the group size hypothesis. But such an increase can arise just through a sample size effect (for example, we expect to roll more sixes the more dice we roll). Furthermore, the increases in probabilities associated with group size, as shown in their Fig. 1, are even less than what would be expected from the sampling size effect. We believe, therefore, that their data fail to support the group size hypothesis.

We show this statistically by comparing the null hypothesis, H_0 : the probability P of an individual drawing a simple artefact on the last step of the experiment is independent of the size of the group to which the individual belongs, with both the one-sided hypothesis that expresses the group size hypothesis, H_{1a} : the probability p of an individual drawing a simple artefact on the last step of the experiment decreases with the size of the group to which the individual belongs, and the alternative one-sided hypothesis that expresses the opposite of the group size hypothesis, namely H_{1b} : the probability p of an individual drawing a simple artefact on the last step of the experiment increases with the size of the group to which the individual belongs.

We first test whether we can reject H_0 . To do this, we determine the expected performance for each group of size $N = 4, 8$ or 16 , using the data for groups of size $N = 2$ to compute a maximum likelihood estimate $\hat{P} = 0.73$ for P , the probability that an individual in a group of size $N = 2$ draws a simple artefact on the last step of the experiment. We then compute for each group of size $N > 2$ the exact probability, using $\hat{P} = 0.73$, that the observed performance for a group of size N deviates from the expected performance under H_0 as much as, or more than, what Derex *et al.*² observe in their experiments.

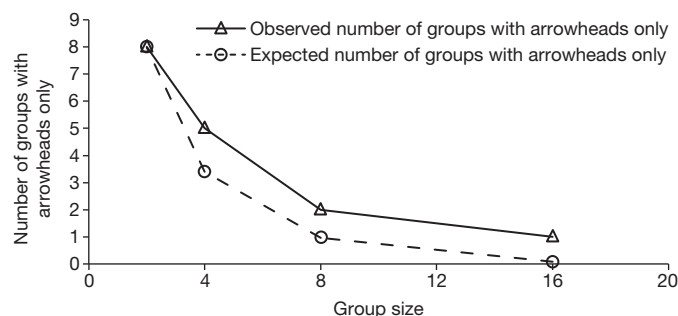


Figure 1 | For each group size, observed and expected number of groups drawing only the simple artefact on the last step of the experiment. Expected values are from a maximum likelihood estimate $\hat{P} = 0.73$ for the probability p that an individual draws a simple artefact on the last step. The expected number (EN) of groups that only draw simple artefacts is given by $EN = \hat{P}^N \times n$, in which N is the group size and n is the number of groups of size N ($n = 15$ for groups of size $N = 2$, otherwise $n = 12$).

The probability P of getting outcomes as extreme as, or more extreme than, the experimental data are given by $P = \Pr(X_4 \geq 5) \Pr(X_8 \geq 2) \Pr(X_{16} \geq 1) = (0.235)(0.252)(0.075) = 0.0045$ (where X_N has a $\text{Bin}(n, 0.73^N)$ distribution, $N = 4, 8$ or 16), as (1) there were five groups of size $N = 4$, two groups of size $N = 8$, and one group of size $N = 16$ that only drew the simple artefact on the last step; and (2) the experiments for each group size are mutually independent. As $P < 0.005$, we reject H_0 .

Next, we note that the direction of deviation from the expected performance under H_0 is that of worse performance (see Fig. 1, in which fewer of the groups with $N > 2$ have at least one individual drawing a complex artefact on the last step of the experiment than would be expected based on the performance of individuals in groups of size $N = 2$). Thus, we also reject H_{1a} and so we conclude that the data support H_{1b} , namely that a larger group size decreases the likelihood of drawing the more complex artefact on the experiment's last step.

In sum, the experimental results of Derex *et al.*² appear to be contrary to the group size hypothesis. This conclusion also affects negatively their other claims and, in general, underscores the importance of identifying the null hypothesis with care.

Methods

We compute, for each group size $N > 2$, $\Pr(X_N \geq x_N)$, in which X_N has a $\text{Bin}(n, 0.73^N)$ distribution, $N = 4, 8$ or 16 , n is the number of groups of size N ($n = 12$ for each of the groups of size $N = 4, 8$, or 16), and x_N is the number of groups of size N that only draw arrowheads on the last step of the experiment.

Claes Andersson¹ & Dwight Read²

¹Physical Resource Theory/Department of Energy and Environment, Chalmers University of Technology, 412 96 Göteborg, Sweden.

²Department of Anthropology, University of California, Los Angeles, Los Angeles, California 90095, USA.

email: dread@anthro.ucla.edu

Received 15 March; accepted 10 April 2014.

- Henrich, J. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *Am. Antiq.* **69**, 197–214 (2004).
- Derex, M., Beugin, M.-P., Godelle, B. & Raymond, M. Experimental evidence for the influence of group size on cultural complexity. *Nature* **503**, 389–391 (2014).
- Collard, M., Kemery, M. & Banks, S. Causes of toolkit variation among hunter-gatherers: a test of four competing hypotheses. *Can. J. Archaeol.* **29**, 1–19 (2005).
- Collard, M., Briggs, B., Ruttle, A. & O'Brien, M. J. Niche construction and the toolkit of hunter-gatherers and food producers. *Biol. Theory* **6**, 251–259 (2011).
- Collard, M., Buchanan, B., O'Brien, M. J. & Scholnick, J. Risk, mobility or population size? Drivers of technological richness among contact-period western North American hunter-gatherers. *Phil. Trans. R. Soc. B* **368**, 20120412 (2013).
- Collard, M., Buchanan, B. & O'Brien, M. J. Population size as an explanation for patterns in the Paleolithic archaeological record: more caution is needed. *Curr. Anthropol.* **54** (S8), S388–S396 (2013).
- Read, D. Tasmanian knowledge and skill: maladaptive imitation or adequate technology? *Am. Antiq.* **71**, 164–184 (2006).
- Read, D. An interaction model for resource implement complexity based on risk and number of annual moves. *Am. Antiq.* **73**, 599–625 (2008).
- Read, D. The misuse of a mathematical model: the Tasmanian case (Reply to Henrich's response). *eScholarship Univ. California* <http://escholarship.org/uc/item/88k8g4rj> (2009).
- Read, D. Population size does not predict artifact complexity: analysis of data from Tasmania, Arctic hunter-gatherers, and Oceania. *eScholarship Univ. California* <http://www.escholarship.org/uc/item/61n4303q> (2012).

Author Contributions D.R. and C.A. contributed equally to this Brief Communications Arising.

Competing Financial Interests Declared none.

doi:10.1038/nature13411

DereX *et al.* replyREPLYING TO C. Andersson & D. Read *Nature* **511**, <http://dx.doi.org/10.1038/nature13411> (2014)

In the accompanying Comment¹, Andersson & Read challenge our results² that group size influences cultural complexity. Using a dual-task computer game, our experiment demonstrated that an increasing group size prevents the loss of cultural traits (simple and complex), promotes their improvements and prevents cultural richness to disappear². Among these various effects, Andersson & Read¹ are questioning the finding that larger groups favour the persistence of the complex trait.

Andersson & Read¹ claim that the expected number of groups that only exploit the simple task is more than expected if individuals in larger groups were behaving as individuals in smaller groups. Thus, they conclude that group size negatively affects the individual ability to exploit the complex task. Instead our data show the opposite pattern: individuals in larger groups tend to be more able to exploit the complex task (Fig. 1), thus supporting the group size hypothesis.

The first issue with the analysis of Andersson & Read¹ is their estimation of the individual probability to exploit the simple task. Indeed, their method is expected to provide a correct estimation only if individuals behave independently of their other group members. As they claim that individuals' ability is affected by group size, this is quite problematic.

The expected number of groups exploiting only arrowheads should be computed from the probability that a single, isolated individual exploits the simple task. This probability is not available in our data, as individuals were always part of a group. To get around this, we used a generalized linear model on individual data to estimate this probability. We obtained a probability ($P = 0.80$) that a single individual exploits the simple task on the last step of the experiment. Using this probability, we computed the probability of getting outcomes as extreme as, or more extreme than, our observed data for each group size: $\Pr(X_2 \geq 8) = 0.873$; $\Pr(X_4 \geq 5) = 0.595$; $\Pr(X_8 \geq 2) = 0.629$ and $\Pr(X_{16} \geq 1) = 0.295$. Thus, for each group size, the expected number of groups that do not exploit the complex task is not significantly different from expected. We then combine P values using Fisher's method³, and obtain an overall P value of 0.79.

Our initial analysis² showed that the probability of maintaining the complex trait within a group is positively affected by group size. Even if explained by sample size effect, this supports the group size hypothesis: sample size effect is expected to be the main mechanism by which group size affects cultural evolution^{4–6}. Sample size effect should promote cultural evolution, unless the individual probabilities to exploit the task decrease drastically with group size. Here our data suggest that the individual probability of exploiting the simple task increases with group size (Fig. 1). In sum, analyses at group and individual levels support the group size hypothesis and are consistent with results from other recent experimental studies^{7,8}.

Culture is a group process that arises as a result of underlying individual-level mechanisms^{9,10}. In order to study cultural evolution, two levels of analysis are therefore workable. The analysis of Andersson & Read¹ illustrates that individuals' behaviours can hardly be deduced from groups behaviours. Thus, each level of analysis can provide specific information. Depending on the question, analyses should be conducted at one level, or both, but conclusions should always be drawn accordingly.

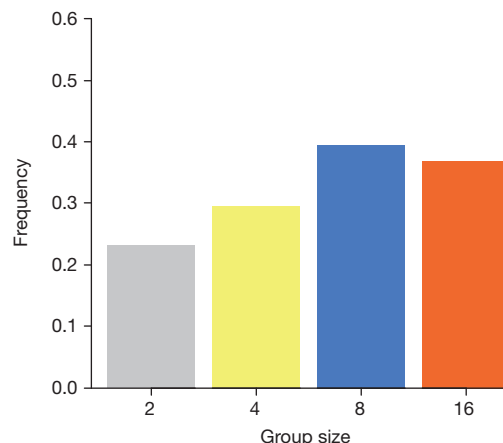


Figure 1 | Frequency of individuals exploiting the complex task according to group size. Group size had a positive and non-significant effect on the individual probability to exploit the complex task (linear: $\chi^2 = 2.53$, d.f. = 1, $P = 0.11$; quadratic: $\chi^2 = 2.15$, d.f. = 1, $P = 0.14$). Older players tend to be less likely to exploit the complex task ($\chi^2 = 3.47$, d.f. = 1, $P = 0.06$). Players who 'died' during the game² were excluded from the analysis (2-player groups: $n = 26$; 4-player groups: $n = 44$; 8-player groups: $n = 86$; 16-player groups: $n = 174$).

Maxime Derex¹, Marie-Pauline Beugin¹, Bernard Godelle¹ & Michel Raymond^{1,2}

¹University of Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

email: maxime.derex@gmail.com

²CNRS, Institute of Evolutionary Sciences, CC 065, Place Eugène Bataillon, Montpellier 34095, France.

- Andersson, C. & Read, D. Group size and cultural complexity. *Nature* **511**, <http://dx.doi.org/10.1038/nature13411> (2014).
- Derex, M., Beugin, M.-P., Godelle, B. & Raymond, M. Experimental evidence for the influence of group size on cultural complexity. *Nature* **503**, 389–391 (2013).
- Fisher, R. A. *Statistical Methods for Research Workers* (Oliver and Boyd, 1925).
- Mesoudi, A. Variable cultural acquisition costs constrain cumulative cultural evolution. *PLoS ONE* **6**, e18239 (2011).
- Henrich, J. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses—the Tasmanian case. *Am. Antiq.* **69**, 197–214 (2004).
- Shennan, S. Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* **11**, 5–16 (2001).
- Muthukrishna, M., Shulman, B. W., Vasilescu, V. & Henrich, J. Sociality influences cultural complexity. *Proc. R. Soc. Lond. B* **281**, 20132511 (2014).
- Kempe, M. & Mesoudi, A. An experimental demonstration of the effect of group size on cultural accumulation. *Evol. Hum. Behav.* <http://dx.doi.org/10.1016/j.jevolhumbehav.2014.02.009> (in the press).
- Richerson, P. J. & Boyd, R. *Not by Genes Alone* (Univ. Chicago Press, 2005).
- Mesoudi, A. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (Univ. Chicago Press, 2011).

doi:10.1038/nature13412

Uncertainties in the timing of unprecedented climates

ARISING FROM C. Mora *et al.* *Nature* **502**, 183–187 (2013)

The question of when the signal of climate change will emerge from the background noise of climate variability—the ‘time of emergence’—is potentially important for adaptation planning. Mora *et al.*¹ presented precise projections of the time of emergence of unprecedented regional climates. However, their methodology produces artificially early dates at which specific regions will permanently experience unprecedented climates and artificially low uncertainty in those dates everywhere. This overconfidence could impair the effectiveness of climate risk management decisions². There is a Reply to this Brief Communication Arising

by Mora, C. *et al.* *Nature* **511**, <http://dx.doi.org/10.1038/nature13524> (2014).

Any human-induced changes in climate will be modulated by natural fluctuations of the oceans and atmosphere (for example, El Niño events). These fluctuations occur randomly and independently, in both reality and individual model-based projections, and act to obscure the climate change signal^{3–5}. Mora *et al.*¹ discuss projections of when changes in climate emerge permanently above the levels of such fluctuations (a metric first considered in ref. 6). However, by ignoring the irreducible limits imposed by these same random fluctuations, Mora *et al.*¹ express their emergence dates with too much certainty.

Several methodological oversights contribute to the erroneous uncertainty quantification. First, Mora *et al.*¹ ignore the possibility that emergence dates before the end of the simulations are not permanent deviations from the historical range⁶ (termed ‘pseudo-emergence’). In many regions where emergence has not occurred by the year 2100, Mora *et al.*¹ artificially set the emergence date to equal 2100. This oversight produces several effects, including: (1) early and overconfident estimates of regional temperature emergence; and (2) implausible emergence dates for precipitation of exactly 2100 with zero uncertainty almost everywhere.

Second, Mora *et al.*¹ estimate precision of regional emergence timing using the standard error of the ensemble mean (σ/\sqrt{N}), where N ($=39$) is the number of simulations and σ is their standard deviation. While the estimate of the ensemble-mean becomes more precise with larger ensemble size, natural fluctuations of the climate (such as El Niño) dictate that the future evolution of climate will not behave like the ensemble mean, but as a single realization from a range of outcomes^{5,7}. The use of σ/\sqrt{N} greatly underestimates⁸ this irreducible uncertainty,

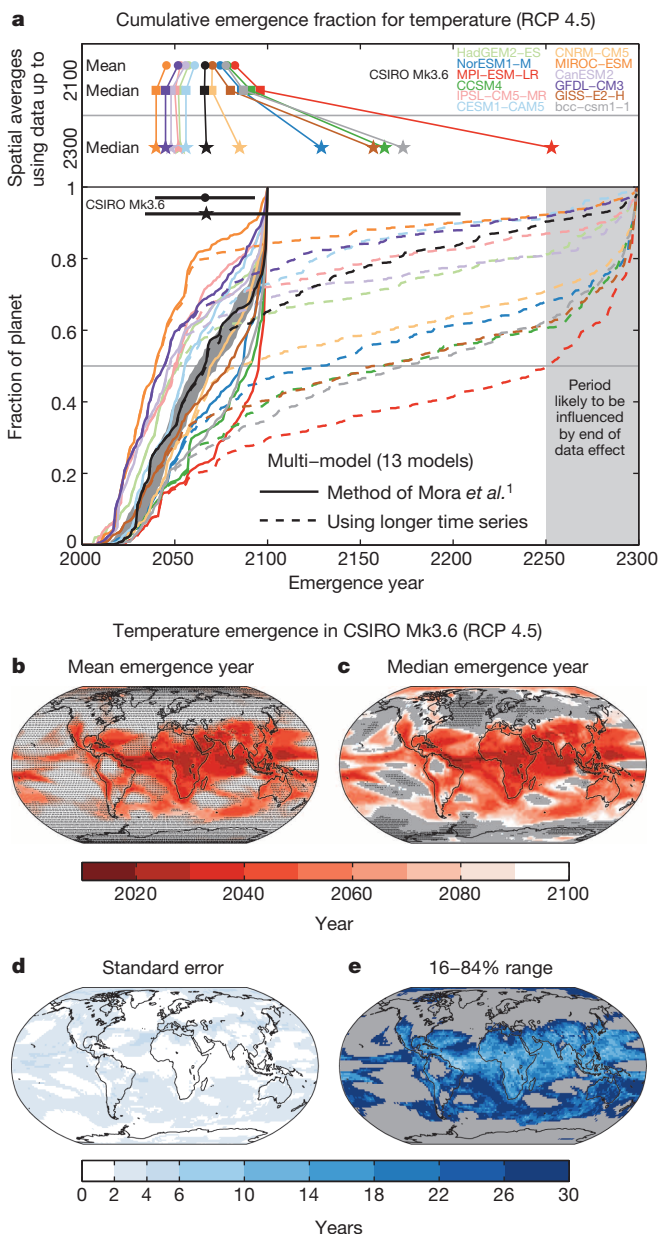


Figure 1 | The year of unprecedented emergence for surface air temperature using RCP4.5. **a**, Lower panel: the cumulative fraction of the planet that has emerged by any particular year for 13 different GCMs when restricting the simulations to end in 2100 (solid lines, as in ref. 1) and in 2300 (dashed lines). The grey shaded region highlights that the end-of-simulation pseudo-emergence effect probably also affects the post-2100 emergence dates after about 2250. For CSIRO Mk3.6 (black curve), spatial variations in the grid-point emergence values are given by the global mean $\pm 1\sigma$ (black circle and bar, as in ref. 1) and a more appropriate 16–84% range of emergence times in which 68% of the grid points lie (black bar and star). The grey shading around the black curve represents the range in coverage for the period 2000–2100 amongst the 30 CSIRO Mk3.6 simulations analysed in b–e. **a**, Upper panel: the year of global emergence using means and data up to 2100 (as in ref. 1, coloured circles), medians and data up to 2100 (coloured squares), and medians and data up to 2300 (coloured stars), showing a substantial delay for several models which do not show median emergence until well after 2100. **b**, The mean (as in the approach of ref. 1) and **c**, median (when considering data up to 2115) emergence year using 30 simulations of CSIRO Mk3.6 GCM. Stippling in **b** shows where at least 1 simulation (of 30) emerges beyond 2100; as such the mean must be estimated by arbitrarily setting post-2100 emergence dates to 2100 (61% by area). Grey regions in **c** show where more than half the simulations emerge after 2100 and hence no median emergence value can be determined (28% by area). Stippling in **c** indicates where all members show emergence beyond 2100 (8% by area). **d**, The standard error about the mean (as in the approach of ref. 1) and **e**, the more appropriate and much larger 16–84% range about the median, with grey regions showing where less than 84% of the simulations have emerged by 2100 and hence no 16–84% range can be estimated (44% by area).

as well as the climate-response uncertainty given by the inter-model spread, and is therefore inappropriate for use in emergence estimates. Given $N = 39$ simulations, there is greater than 85% chance that the actual emergence time at any location will fall outside the uncertainty values of ref. 1, and an infinite number of simulations would implausibly suggest zero uncertainty in the projected emergence time. Nor can the standard error simply be scaled to a more appropriate uncertainty range (for example, a 16%–84% range, equivalent to $\pm 1\sigma$ for a normal distribution), partly because the ‘right-censored’ emergence results of ref. 1 have an explicit upper-bound of 2100, making their distribution highly non-normal.

To demonstrate the impact of these oversights, we have replicated the analysis of ref. 1 for surface air temperature using: (1) a multi-model ensemble of simulations that extends to the year 2300; and (2) a large ensemble of simulations from a single model that extends to 2115.

Mora *et al.*¹ report that their “index has a global mean of 2069 (± 18 years s.d.) for near-surface air temperature” in the RCP4.5 forcing pathway, where “s.d.” refers to the spatial standard deviation of their grid-point means. However, the end-of-simulation effects invalidate the concept of a global mean permanent emergence date. Even apparent emergence years as early as 2050 may not actually permanently emerge until post-2100, as evidenced by the time of divergence between dashed and solid lines in Fig. 1a. Further, 41% (multi-model median) of the pre-2100 emergence values (by area) are either pseudo-emergent (31%) or artificially set to 2100 (10%). We also find that no model shows permanent emergence everywhere by 2100, or even by 2250. The large fraction of grid points exhibiting post-2100 emergence also highlights the underestimation of spatial emergence variability: whereas Mora *et al.*¹ report a spatial s.d. of ± 18 years, the 16–84% grid-point range is >150 years for virtually all of the models (see specific example for CSIRO Mk3.6 model in Fig. 1a). Finally, while some global-median emergence estimates using post-2100 data are similar to the global-mean (and global-median) estimates using only pre-2100 data, such agreement is fortuitous—as evidenced by the substantial delay in several models (compare coloured circles and stars in Fig. 1a)—and should not be expected *a priori* for the multi-model mean values from ref. 1.

The large single-model ensemble helps clarify the spatial pattern of irreducible uncertainty (Fig. 1b). In this ensemble, 61% of the planet exhibits the possibility of post-2100 emergence, thwarting the calculation of mean emergence and biological impacts in these regions (including Amazonia and the Southern Ocean which are in the biodiversity hotspots of ref. 1). In addition, the standard errors (as used by Mora *et al.*¹) are less than 6 years everywhere, whereas the irreducible 16–84% uncertainty range is more than 6 years everywhere, and 75% of the planet has a 16–84% range of more than 20 years. Note that inter-model uncertainty will further increase the spread in grid-point emergence times (compare the multi-model spread with the shaded intra-model spread in Fig. 1a), and decrease the coverage of well-defined grid-point averages. Finally, while the delay in emergence and increase in uncertainty is evident for annual temperatures (the primary metric of ref. 1), it will be more pronounced for other variables analysed¹, such as monthly temperatures and precipitation, but less pronounced for annual temperatures in higher forcing pathways.

Last, the main conclusion of ref. 1—of early tropical emergence—is already a key summary statement in the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5): “Relative to natural internal variability, near-term increases in seasonal mean and annual mean temperatures are expected to be larger in the tropics and subtropics than in mid-latitudes (high confidence)” (ref. 9). The reason for ‘high confidence’ is that tropical temperature emergence has already been seen in observations^{6,10–12} and in many previous studies examining climate simulations^{3,4,6,11,13–18}, none of which were cited by ref. 1. While projections of emergence times are clearly important for estimating a

wide range of impacts (as demonstrated for food security¹⁷, biodiversity hotspots¹⁸ and ocean biogeochemistry¹⁹), they need to be quantified within a framework that incorporates climate variability, as illustrated in the large body of literature that has already examined this issue.

METHODS

We use simulations of surface air temperature from 13 global climate models (GCMs) given historical radiative forcings from 1860–2005 and the RCP4.5 forcing pathway from 2006–2300. We estimate the unprecedented emergence time for every grid point in each simulation independently. The cumulative fraction of emergence (Fig. 1a) shows the proportion of the surface area of the planet that has emerged by each year. The emergence calculations are repeated while restricting the data to end in either 2100 or 2300. We also use an ensemble of 30 simulations of the CSIRO Mk3.6 GCM²⁰ which were given the same radiative forcings for the period 1860–2115. The CSIRO GCM is chosen because of the availability of the large ensemble of simulations and its similarity to the multi-model mean/median behaviour. This ensemble of emergence years is used to estimate the grid-point averages and uncertainty ranges (Figs 1b–e).

Ed Hawkins¹, Bruce Anderson², Noah Diffenbaugh³, Irina Mahlstein⁴, Richard Betts^{5,6}, Gabi Hegerl⁷, Manoj Joshi⁸, Reto Knutti⁹, Doug McNeall⁵, Susan Solomon¹⁰, Rowan Sutton¹, Jozef Syktus¹¹ & Gabriel Vecchi¹²

¹National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading RG6 6BB, UK.

e-mail: e.hawkins@reading.ac.uk

²Department of Earth and Environment, Boston University, Massachusetts 02215, USA.

³School of Earth Sciences and Woods Institute for the Environment, Stanford University, California 94305, USA.

⁴Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich CH-8058, Switzerland.

⁵Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK.

⁶Department of Geography, Exeter University, Exeter EX4 4RJ, UK.

⁷School of Geosciences, University of Edinburgh, Edinburgh EH9 3JW, UK.

⁸Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

⁹Institute for Atmospheric and Climate Science, ETH Zurich, CH-8092 Zurich, Switzerland.

¹⁰Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

¹¹Department of Science, Information Technology, Innovation and the Arts, and University of Queensland, Brisbane, Queensland 4001, Australia.

¹²Geophysical Fluids Dynamics Laboratory, Princeton, New Jersey 08540, USA.

Received 28 October 2013; accepted 29 April 2014.

1. Mora, C. *et al.* The projected timing of climate departure from recent variability. *Nature* **502**, 183–187 (2013).
2. Lempert, R., Nakicenovic, N., Sarewitz, D. & Schlesinger, M. Characterizing climate-change uncertainties for decision-makers. *Clim. Change* **65**, 1–9 (2004).
3. Mahlstein, I., Knutti, R., Solomon, S. & Portmann, R. W. Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.* **6**, 034009 (2011).
4. Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys. Res. Lett.* **39**, L01702 (2012).
5. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nature Clim. Change* **2**, 775–779 (2012); erratum **2**, 888 (2012).
6. Diffenbaugh, N. S. & Scherer, M. Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Clim. Change* **107**, 615–624 (2011).
7. Knutti, R., Furrer, R., Tebaldi, C., Cernak, J. & Meehl, G. A. Challenges in combining projections from multiple climate models. *J. Clim.* **23**, 2739–2758 (2010).
8. Nagele, P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br. J. Anaesth.* **90**, 514–516 (2003).

9. Stocker, T. F. *et al.* (eds) Summary for policymakers. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
10. Ho, C. K., Hawkins, E., Shaffrey, L. & Underwood, F. M. Statistical decadal predictions for sea surface temperatures: a benchmark for dynamical GCM predictions. *Clim. Dyn.* **41**, 917–935 (2013).
11. Anderson, B. T. Near-term increase in frequency of seasonal temperature extremes prior to the 2°C global warming target. *Clim. Change* **108**, 581–589 (2011).
12. Mahlstein, I., Hegerl, G. & Solomon, S. Emerging local warming signals in observational data. *Geophys. Res. Lett.* **39**, L21711 (2012).
13. Hegerl, G. C. *et al.* Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *J. Clim.* **9**, 2281–2306 (1996).
14. Syktus, J. *et al.* Latitudinal dependence of signal-to-noise patterns from two general circulation models with CO₂ forcing. *Clim. Dyn.* **13**, 293–302 (1997).
15. Diffenbaugh, N. S. & Giorgi, F. Climate change hotspots in the CMIP5 global climate model ensemble. *Clim. Change* **114**, 813–822 (2012).
16. Mahlstein, I., Daniel, J. S. & Solomon, S. Pace of shifts in climate regions increases with global temperature. *Nature Clim. Change* **3**, 739–743 (2013).
17. Battisti, D. S. & Naylor, R. L. Historical warnings of future food insecurity with unprecedented seasonal heat. *Science* **323**, 240–244 (2009).
18. Beaumont, L. J. *et al.* Impacts of climate change on the world's most exceptional ecoregions. *Proc. Natl Acad. Sci. USA* **108**, 2306–2311 (2011).
19. Keller, K. M., Joos, F. & Raible, C. C. Time of emergence of trends in ocean biogeochemistry. *Biogeosci. Discuss.* **10**, 18065–18092 (2013).
20. Jeffrey, S. *et al.* Australia's CMIP5 submission using the CSIRO-Mk3.6 model. *Austr. Meteorol. Oceanogr. J.* **63**, 1–13 (2013).

Author Contributions E.H. led the analysis and writing of the manuscript. B.A., N.D. and I.M. performed additional analyses, discussed the results and contributed text. All other authors contributed text.

Competing Financial Interests Declared none.

doi:10.1038/nature13523

Mora *et al.* reply

REPLYING TO E. Hawkins *et al.* *Nature* **511**, <http://dx.doi.org/10.1038/nature13523> (2014)

In the accompanying Comment, Hawkins *et al.*¹ suggest that our index² of the projected timing of climate departure from recent variability is biased to occur too early and is given with overestimated confidence. We contest their assertions and maintain that our findings are conservative and remain unaltered in light of their analysis.

We presented an index² that quantifies the year after which the climate continuously exceeds the bounds of historical variability, using 39 CMIP5 Earth System Models. Uncertainty in climate projections from these models arises chiefly from natural ('internal') climate variability, model error and uncertainty in the future evolution of greenhouse gas concentrations³. Hawkins *et al.*¹ suggest that by "ignoring the irreducible limits imposed by... random fluctuations, Mora *et al.*² express their emergence dates with too much certainty". However, our index was calculated independently for individual model simulations, which include internal variability. By considering individual simulations (internal variability) from each of 39 models (model-to-model error), under two emission pathways (scenario uncertainty), our results account for the three major sources of uncertainty in climate projections.

Our analysis of climate departure included only projections to the year 2100 due to their greater availability (only a third of the CMIP5 models have projections beyond 2100). Hawkins *et al.*¹ assert that the use of model projections to the year 2100 reduces the global mean timing of climate departure because we assigned the year 2100 to cells where unprecedented climates might occur after 2100. This is a valid constraint, and thus, climate departure at 2100 in our results should be interpreted as emergence that will occur in 2100 or later, or not at all.

The implications of using projections to 2100 are, however, exaggerated by Hawkins *et al.*¹. First, we recalculated our index using the multi-model median, which is less affected than the mean by outlier projections of climate departure after 2100, and found only small differences. The global median temperature departure was 2076 under RCP4.5 (the reported global mean was 2069) and 2045 under RCP8.5 (the reported global mean was 2047). The multi-model median delivers similar results to the mean, even if projections to 2300 are used. For instance, the analysis in ref. 1 shows that 7 out of 13 models exhibit similar or even earlier global median temperature departures using projections to 2300 compared to global mean based on projections to 2100 (upper plot in figure 1a in ref. 1). Second, Hawkins *et al.*¹ chose RCP4.5, stating that the limitation is "less pronounced for annual temperatures in higher forcing pathways". Indeed, by 2080, 97% of the planet will face temperature departure for the remainder of the twenty-first century

under RCP8.5. Under the RCP4.5 pathway, 67% of the planet will face temperature departure by 2080, highlighting the imminent departure of Earth's climate even under an optimistic mitigation scenario. Finally, from a global biological and social perspective, the potential limitation associated with climate departures beyond 2100 is small, as it is relevant only to high latitudes and not for areas where the majority of people and species on the planet live.

Any statistical value should be interpreted based on the metric it represents. Hawkins *et al.*¹ claim that by reporting the standard error of the mean our results are given with too much confidence. Our paper is transparent and clearly states that the standard error of the mean was our metric of uncertainty among models, and although the values provided should be interpreted in the context of that metric, they can easily be converted to another choice of statistic if so desired (for example, our standard error can be multiplied by \sqrt{N} to obtain the standard deviation). Hawkins *et al.*¹ further suggest that the standard error is the wrong choice as "the future evolution of climate will not behave like the mean, but as a single realization from a range of outcomes". In other words, all models' simulations are equally likely, and thus, statistics that describe the broad range of projections are more suitable. This premise, however, conflicts with findings that it is the multi-model average that best approximates mean observed conditions, often better than any individual model, as demonstrated by prior studies⁴ and confirmed in our paper². Although there is no established "correct" way to express uncertainty, at least for the results from Earth System Models that can be verified, metrics of variability around the consensus mean are more appropriate.

Hawkins *et al.*¹ also suggested that the standard deviation cannot be scaled to their suggested 16–84% range multi-model dispersion as the climate departures are not normally distributed. However, "contrary to popular misconception, the standard deviation is a valid measure of variability regardless of the distribution. About 95% of observations on any distribution usually fall within 2 standard deviation limits..."⁵. We recalculated multi-model uncertainty as the standard deviation and as the 16–84% range among model projections for temperature and found small differences. The global median multi-model uncertainty estimates by these two metrics differ by 2.5 years under RCP4.5 and 1.6 years under RCP8.5.

Our paper² used all (not a subset) of the latest generation of Earth System Models that have complete projections for RCP4.5, RCP8.5 and historical experiments, under very conservative criteria for estimating climate departure (for example, using the minimum and maximum

historical values to set bounds, defining climate departure as the year after which all subsequent years are out of historical bounds, and using a historical period already affected by human greenhouse gas emissions; we demonstrated that all these criteria delay the estimated year of climate departure). We also used data on species distributions, protected areas and socio-economic conditions to show that the earliest emergence of unprecedented climates will occur in areas with the greatest number of species on Earth, where a large proportion of the world's human population lives and where conservation and economic capacity to adapt are limited. These conclusions remain unaltered.

Camilo Mora¹, Abby G. Frazier¹, Ryan J. Longman¹, Rachel S. Dacks², Maya M. Walton^{2,3}, Eric J. Tong^{3,4}, Joseph J. Sanchez¹, Lauren R. Kaiser¹, Yuko O. Stender^{1,3}, James M. Anderson^{2,3}, Christine M. Ambrosino^{2,3}, Iria Fernandez-Silva^{3,5}, Louise M. Giuseffi¹ & Thomas W. Giambelluca¹

¹Department of Geography, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, USA.

e-mail: cmora@hawaii.edu

²Department of Biology, University of Hawaii at Mānoa, Honolulu, Hawaii 96822, USA.

³Hawaii Institute of Marine Biology, University of Hawaii at Mānoa, Kāne'ohe, Hawaii 96744, USA.

⁴Department of Oceanography, University of Hawaii at Mānoa, Honolulu, Hawaii 96822, USA.

⁵Trans-disciplinary Organization for Subtropical Island Studies (TRO-SIS), University of the Ryukyus, Senbaru, Nishihara, Okinawa 903-0213, Japan.

1. Hawkins, E. *et al.* Uncertainties in the timing of unprecedented climates. *Nature* **511**, <http://dx.doi.org/10.1038/nature13523> (2014).
2. Mora, C. *et al.* The projected timing of climate departure from recent variability. *Nature* **502**, 183–187 (2013).
3. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nature Clim. Change* **2**, 775–779 (2012).
4. Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A* **365**, 2053–2075 (2007).
5. Altman, D. G. & Bland, J. M. Standard deviations and standard errors. *Br. Med. J.* **331**, 903 (2005).

doi:10.1038/nature13524

Use it or lose it

A comparison of the development of adapted limbs in mammals uncovers multiple avenues to digit loss and highlights that early patterning events are not untouchable by evolutionary tinkering. [SEE ARTICLES P.41 & P.46](#)

BAU-LIN HUANG & SUSAN MACKEM

Limb modifications are fundamental to tetrapod evolution. Adaptive digit loss enables specialized functions such as running or flight, and has repeatedly evolved in parallel. But the developmental mechanisms underlying deviation from the five-digit ground state are unclear, partly owing to the hurdles involved in analysing embryos from animals that are not typically studied in the laboratory. Two papers published in this issue, from Cooper *et al.*¹ (page 41) and Lopez-Rios *et al.*² (page 46), illuminate developmental mechanisms that result in digit reduction in hoofed mammals and in rodents, and reveal surprising plasticity in the pathways that lead to evolutionary convergence. Together, these studies show that similar developmental changes may be shared by distantly related groups, but also that fundamentally different mechanisms can be used in closely related taxa to yield similar digit-loss outcomes.

During embryonic development, interacting and cross-regulating signalling centres in the early limb bud control and coordinate digit patterning (type and number) and growth³. Among these, the apical ectodermal ridge (AER) region secretes fibroblast growth factors (FGFs) that direct outgrowth, and the posterior limb bud secretes Sonic hedgehog (SHH), which regulates digit patterning dose-dependently through its posteriorly polarized distribution (Fig. 1a). Digit formation fails without SHH, but precocious SHH activation also causes digit loss by disrupting polarity and signalling-centre formation⁴. After digit precursors form, FGFs in the AER stimulate digit-tip elongation⁵, and bone morphogenetic proteins (BMPs) reshape limb borders and free individual digits through apoptotic cell death⁶, and modulate growth by inducing AER regression⁷.

Cooper *et al.* compared the expression of key patterning genes and remodelling activity in limbs of embryonic ungulates (hoofed mammals with 1–4 toes) and rodents (particularly jerboas, hopping desert rodents with 3 or 5 toes) — two groups that diverged around 60 million years ago (Fig. 1b). No early patterning changes distinguished a 3-toed (*Dipus sagitta*) from a 5-toed (*Allactaga elater*)

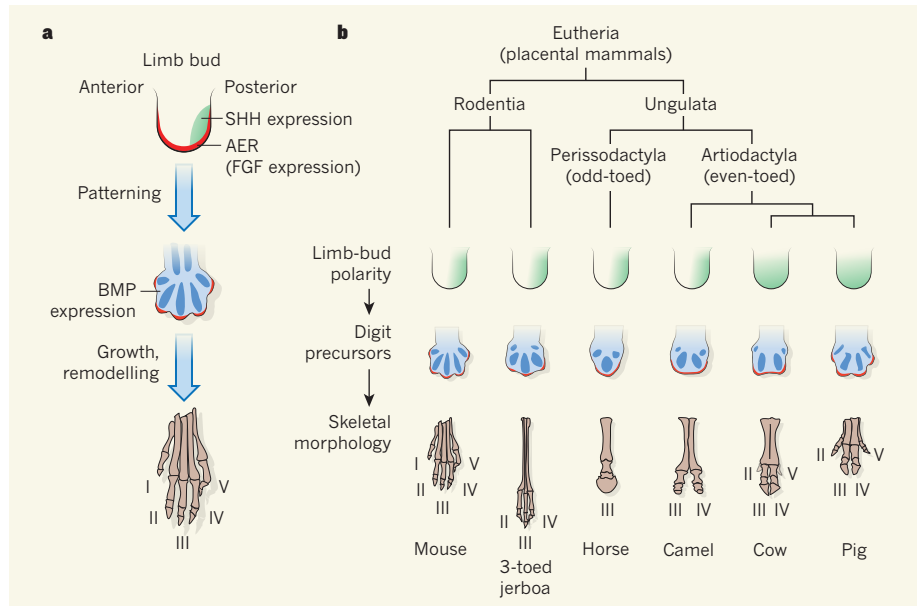


Figure 1 | Evolution of digit loss. **a**, In the developing limb bud, polarity in the expression of the signalling molecule SHH and the action of fibroblast growth factors (FGFs) in the apical ectodermal ridge (AER) control early limb-bud patterning and outgrowth, forming a digit blueprint. Signalling through the broadly expressed bone morphogenetic proteins (BMPs) then regulates remodelling processes, including cell death and proliferation (by causing the AER to regress), to determine final skeletal morphology. **b**, Cooper *et al.*¹ and Lopez-Rios *et al.*² find that mechanisms of digit loss, involving both early limb-bud patterning and later remodelling and growth of digit precursors, vary across species. In some cases, the same mechanisms are shared by groups that are widely separated evolutionarily, such as jerboas and camels. In other cases, different mechanisms are used within more closely related groups, such as horses and camels compared with cows and pigs. (Adult cows retain highly reduced digits II and V as two vestigial dew claws. Late-stage pig limb-bud data are taken from ref. 12.)

jerboa species, or from the 5-toed mouse and 5-fingered forelimbs in all three rodents. Instead, in 3-toed jerboas, expanded apoptotic regions encompassed the digit I and V precursors and distal proliferative tips. Examination of the pro-apoptotic BMP pathway implicated increased expression of *Msx2*, which encodes a target transcription factor of the BMP pathway, as the triggering event. Secondary premature AER regression over digit tips also contributed to digit loss. Proof that *MSX2* is indeed the primary instigator of adaptive digit loss in jerboas awaits genome comparisons and experimental interrogation of gene-regulatory regions in mice.

Extending their analysis to two different ungulate groups — the even-toed artiodactyls (including 2-toed camels and 4-toed pigs) and the odd-toed (1-toed) horses — Cooper

et al. found that camels and horses also use cell-death mechanisms to reduce digit number. However, although apoptosis is the major factor in promoting digit loss in camels, the authors' findings suggest that *MSX2* is not the instigator. In horses, although digit II and IV precursors are lost through cell death, digit I and V precursors never form, suggesting an additional patterning change that warrants further analysis. Nevertheless, these results point to cell death and growth arrest as major drivers of convergent digit loss in mammalian evolution. It will be interesting to see if progressive digit loss in other vertebrates, such as birds⁷, is similarly driven.

Unexpectedly, Cooper *et al.* found that the cell-death machinery is unaltered in some ungulates. In pigs, expression of limb-bud patterning genes loses polarity, with reduced

expression of *Ptch1*, which encodes the SHH receptor. When unbound, PTCH1 suppresses SHH-pathway activity; SHH binding derepresses the pathway and upregulates *Ptch1* expression, which sequesters SHH, forming a negative feedback loop³. Lopez-Rios *et al.* also found reduced *Ptch1* expression in cows — the focus of their work — that was associated with non-polarized expression of other SHH targets and broader spread of SHH protein across the limb bud. Furthermore, they showed that the limb buds of mouse embryos with limb deletion of *Ptch1* lose polarity and develop artiodactyl-like digits, implicating reduced *Ptch1* as the primary basis for digit loss in cows.

Lopez-Rios and colleagues then used cutting-edge analysis of dynamic chromatin topology to assess interactions between gene-regulatory and promoter regions. This identified a mouse *Ptch1* limb-regulatory module (LRM) that drove expression of a reporter gene mimicking *Ptch1* in the limb. By contrast, the cow *Ptch1* LRM was unable to drive reporter expression appropriately. Intriguingly, the LRM contains artiodactyl-specific DNA insertions that disrupt a highly evolutionarily conserved region. However, this region by itself failed to regulate reporter-gene expression appropriately, and the disrupting insertions are present in camels, which express *Ptch1* normally. The DNA changes that restore or ablate the function of bovine and mouse LRMs, respectively, remain to be identified — this will be necessary to test whether the LRM change suffices for loss of bovine *Ptch1* activation.

These papers show how independent molecular mechanisms can drive evolutionary convergence. It is likely that multiple genetic hits involving *Ptch1* regulation, changes in BMP–MSX2-pathway regulation and other unknown genetic changes, separately and in combination, yielded digit loss in ungulates. Although cows and pigs share reduced *Ptch1* expression, their extent of digit reduction differs. In addition, a shift in weight-bearing from middle-toe-centred to symmetric sharing between digits III and IV, the hallmark of even-toed artiodactyls, was already present in their pentadactyl ancestors and persists in camels — without obvious *Ptch1* changes. Consequently, these artiodactyl features are either unrelated to *Ptch1* expression or reflect progressive, incremental changes in *Ptch1* regulation. Dissecting the diverse mechanisms converging on similar structures will require characterization of the tissue-specific regulation for each candidate gene. Mutations to regulatory regions are suspected to be major drivers of evolution, but few vertebrate examples have been characterized^{8,9}. Lopez-Rios and colleagues' study is a benchmark for future work in this area.

It is not surprising that the SHH pathway, which is so central to digit specification, is a prime target for digit adaptation. In different Australian skink species, which have between two and five digits, digit loss correlates with the

time span of *Shh* expression at late limb-bud stages¹⁰, a finding that has been functionally corroborated by genetic studies in mice¹¹. But digit loss due to late SHH loss does not alter morphologies and is more related to growth arrest than patterning. Lopez-Rios and colleagues' work shows that early, precocious SHH activity can also be fine-tuned to create a symmetrical, but functional cow foot. Intriguingly, both patterning and remodelling mechanisms also lead to premature AER regression and FGF loss, which points to FGF modulation as another major driver of evolutionary limb adaptation. These mechanisms are linked by a common set of interacting signals, whose early and late developmental roles may often collaborate to achieve a final adaptive morphology.

The high incidence of convergent adaptive evolution of the limb derives from its intricate and robust — but plastic — regulatory network. Developmental mechanisms have long been considered the major constraint on evolutionary change. However, these studies invite speculation that, in response to selective pressures to adapt, the 'robustness' of the

regulatory network guiding limb development evolved to be evolvable. ■

Bau-lin Huang and Susan Mackem are in the Cancer and Developmental Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702, USA.
e-mail: mackems@mail.nih.gov

1. Cooper, K. L. *et al.* *Nature* **511**, 41–45 (2014).
2. Lopez-Rios, J. *et al.* *Nature* **511**, 46–51 (2014).
3. Butterfield, N. C., McGlinn, E. & Wicking, C. *Curr. Top. Dev. Biol.* **90**, 319–341 (2010).
4. Zhulyn, O. *et al.* *Dev. Cell* **29**, 241–249 (2014).
5. Sanz-Ezquerro, J. J. & Tickle, C. *Curr. Biol.* **13**, 1830–1836 (2003).
6. Zuzarte-Luis, V. & Hurler, J. M. *Semin. Cell Dev. Biol.* **16**, 261–269 (2005).
7. de Bakker, M. A. *et al.* *Nature* **500**, 445–448 (2013).
8. Cretekos, C. J. *et al.* *Genes Dev.* **22**, 141–151 (2008).
9. Chan, Y. F. *et al.* *Science* **327**, 302–305 (2010).
10. Shapiro, M. D., Hanken, J. & Rosenthal, N. J. *Exp. Zool.* **B 297**, 48–56 (2003).
11. Zhu, J. *et al.* *Dev. Cell* **14**, 624–632 (2008).
12. Sears, K. E. *et al.* *Evol. Dev.* **13**, 533–541 (2011).

This article was published online on 18 June 2014.

ASTROPHYSICS

A tight duo in a trio of black holes

Tight pairs of supermassive black holes are expected to emit gravitational waves that could give astronomers a new way to explore the cosmos. One relatively tight pair has been discovered within a rare triple system. SEE LETTER P.57

GREG TAYLOR

Observations have demonstrated that every galaxy with a mass comparable to that of the Milky Way or larger hosts at its centre a supermassive black hole — millions to billions of times more massive than the Sun. These large galaxies are thought to form by mergers of two (or more) smaller galaxies, so the natural expectation is that astronomers should frequently observe pairs of supermassive black holes. Such binary black hole systems are of great interest because they could have a profound influence on the centre of the galaxy, stirring up its gas and ejecting its stars. However, despite many searches^{1,2}, so far we know of only a few binary black hole systems. On page 57 of this issue, Deane *et al.*³ present the discovery of a rare triple system, and report an improved method for finding tight pairs.

As they coalesce, tight binary supermassive black hole systems emit powerful gravitational waves — ripples in the fabric of space-time that propagate at the speed of light. The detection

of these waves would provide additional confirmation of Einstein's general theory of relativity, and would give astrophysicists a new way to explore the cosmos. These waves may be detectable with future space-based instruments⁴, which would work by bouncing lasers off widely separated spacecraft and looking for subtle changes in the measured distance between them due to a passing gravitational wave. Another detection method under active development is to use ground-based pulsar timing arrays⁵, such as NANOGrav⁶. This method takes advantage of the wonderful regularity of pulsating stars by carefully timing the arrival of pulses from many such stars and then looking for tiny changes in the arrival times induced by passing gravitational waves.

Many searches have been carried out for tight binary supermassive black holes. These searches have been made at wavebands from radio to X-rays, because supermassive black holes strongly emit across the electromagnetic spectrum. Radio observations are particularly interesting because the technique of very long baseline interferometry (VLBI)



50 Years Ago

“In spite of all our deficiencies in knowledge, I believe we are now in prospect of a comprehensive synthesis in understanding of the means of action of chemical, physical, viral and indeed purely genetic carcinogenic influences.” These brave words were written recently by Prof. Alexander Haddow in his introduction to a recent issue of the *British Medical Bulletin*. Similar statements in the past have been premature, but the thirteen review articles that he is introducing give Prof. Haddow some justification for his optimism. He feels that in time these processes can be elucidated with chemical precision to reveal the differences, on the cytogenetic and cellular levels, between normal and cancer cells.

From *Nature* 4 July 1964

100 Years Ago

In reading through Mr. Cunningham's article, I have been struck by several points which seem to require some elucidation. ... Whether these difficulties are inherent in the “principle of relativity” as it is generally understood, or whether it is merely owing to the fact that for some time I have been accustomed to look at the matter from a different point of view, which I believe solves these difficulties, I cannot say.

A.A. ROBB.

It should be fairly clear that the articles referred to by Mr. Robb were written with the intention of showing the need for a revision of the common ideas about space and time. ... But the articles did not profess to describe in detail a logical scheme of the universe of motion. Mr. Robb's forthcoming work in which this is attempted is anticipated with much interest.

E. C.

From *Nature* 2 July 1914

— which creates a giant radio telescope, thousands of kilometres across, from a collection of smaller telescopes — provides the highest spatial resolution achievable. Such high resolution allows the examination of black hole separations down to parsec scales (1 parsec is 3×10^{16} metres), even in the most distant galaxies. The Caltech–Jodrell Bank VLBI survey, initiated in the 1990s, led to the discovery⁷ of the tightest binary black hole system found so far, which is known to astronomers as 0402 + 379 and has a separation of just 7 pc (Fig. 1). Additional VLBI surveys examined ten times more radio sources than the Caltech–Jodrell Bank VLBI survey, but no further tight binary black hole systems were found.

The crucial issue for finding tight binaries is how quickly the black hole merger takes place. The initial decay (shrinkage) of the orbits of the binary supermassive black holes probably happens quickly thanks to frequent gravitational interactions between the two black holes and stars⁸. At a black hole separation of under 10 pc, the density of stars drops off, and the binary system could ‘stall’ before gravitational radiation can begin to carry away angular momentum and allow the orbital decay to resume. If this stalling phase exists, then tight binary black holes should be commonplace. However, if gas or stellar interactions can bridge the gap to sub-parsec scales⁹ at which gravitational radiation becomes efficient at shrinking the orbits, then tight binary black hole systems could be truly rare.

The triple black hole system discovered by Deane and colleagues, known as J1502 + 1115, was initially identified¹⁰ as a quasar (a mass-accreting supermassive black hole system) with double-peaked optical emission lines. Such double-peaked lines have been suggested as a sign of binary black hole systems, in which each of the black holes has its own peak and the two peaks are shifted, with respect to the systematic velocity of the host galaxy, by their orbital velocity. Unfortunately, many single supermassive black hole systems have complicated mass-accretion disks around them that can also produce two (or more) emission-line peaks¹¹.

To confirm a candidate system, other evidence such as multiple compact radio jets is needed. And in J1502 + 1115 there is a bonanza of three such compact jets, with two of them separated by just 140 pc, as Deane *et al.* now report using VLBI observations. Furthermore, the authors find that, in this system, the large-scale radio emission associated with the tight pair has an ‘S’-symmetry suggestive of precession of the jets such as might be expected in a binary system. Such a symmetry could provide a way to search for other tight pairs. The authors convincingly argue against other possible explanations for the observed radio structures (for example, the possibility of a double jet from a single black hole). What is more, they suggest that tight binary black holes are actually more commonplace

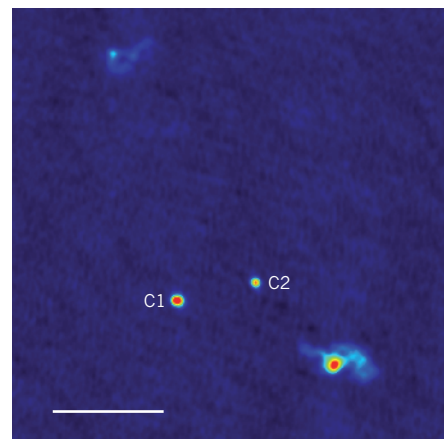


Figure 1 | A tight binary. A radio image of the binary supermassive black hole system 0402 + 379, taken at a frequency of 22 gigahertz with the Very Long Baseline Array. It shows the two black holes (C1 and C2) between extended radio jets, which straddle C2. These jets are launched by an unseen matter-accretion disk around C2, and are observed to be expanding⁷. Deane *et al.*³ have detected a triple black hole system that has a tight pair separated by about 140 parsecs. Colours indicate strength of the radio emission: red shows strong emission and blue weak. Scale bar, 10 pc.

than previously thought, on the basis of their success in finding one new system out of six candidate double-peaked systems that have moderate radio emission and host-galaxy brightness.

If tight binary black holes actually are long-lived and therefore commonplace, we should not have long to wait for many more of them to be found. Many candidate systems already fit Deane and colleagues' observational selection criteria, and new optical surveys using the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) and the Large Synoptic Survey Telescope (LSST) will greatly increase the numbers of known active galaxies containing at least one supermassive black hole.

Our best evidence for tight systems relies heavily on high-resolution VLBI observations. However, some care is warranted because jets can exhibit a wide variety of morphologies and spectra, and are thus subject to misinterpretation. The ultimate test for a tight binary black hole system would be to observe the orbital motions of the two components directly. Such a feat would require decades of monitoring with VLBI even for the 0402 + 379 system, which has a predicted orbital period of 150,000 years, but it could be accomplished in just a few years if an even tighter system can be found. ■

Greg Taylor is in the Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA.
e-mail: gbtaylor@unm.edu

1. Helmboldt, J. F. *et al.* *Astrophys. J.* **658**, 203–216 (2007).
2. Condon, J., Darling, J., Kovalev, Y. Y. & Petrov, L. *Proc. 11th Asian-Pacific Regional IAU Meet., NARIT Conf.*

Ser. Preprint at <http://arXiv.org/abs/1110.6252> (2011).

3. Deane, R. P. *et al. Nature* **511**, 57–60 (2014).
4. Hughes, S. A. *Ann. Phys.* **303**, 142–178 (2003).
5. Detweiler, S. *Astrophys. J.* **234**, 1100–1104 (1979).
6. <http://nanograv.org/>
7. Rodriguez, C. *et al. Astrophys. J.* **646**, 49–60 (2006).
8. Begelman, M. C., Blandford, R. D. & Rees, M. J.

- Nature* **287**, 307–309 (1980).
9. Milosavljević, M. & Merritt, D. *Astrophys. J.* **563**, 34–62 (2001).
10. Smith, K. L. *et al. Astrophys. J.* **716**, 866–877 (2010).
11. Gaskell, C. M. *Nature* **463**, E1–E2 (2010).

This article was published online on 25 June 2014.

STRUCTURAL BIOLOGY

Lipopolysaccharide rolls out the barrel

Two crystal structures of the LptD–LptE protein complex reveal how the cell-wall component lipopolysaccharide is delivered and inserted into the external leaflet of the bacterial outer membrane. [SEE ARTICLE P.52 & LETTER P.108](#)

RUSSELL E. BISHOP

The inner membrane of many bacteria is surrounded by a protective cell wall, which includes an outer membrane adorned with a tightly packed layer of lipopolysaccharide (LPS) molecules. LPS is an elaborate phospholipid that is extensively decorated with sugars, and its presence provides a barrier against harmful agents such as detergents and lipophilic ('greasy') antibiotics. Many of the cellular processes by which LPS is assembled and transported to the outer membrane are essential for survival, and thus are targets for the development of new antibiotics. In two papers published in this issue, Dong *et al.*¹ (page 52) and Qiao *et al.*² (page 108) present structural studies that disclose a remarkable molecular mechanism by which LPS arrives at its final destination in the outer membrane.

LPS molecules are mainly assembled by biosynthetic enzymes in the cytoplasm or at the cytoplasmic face of the inner membrane, but the activity of several LPS-modifying enzymes that function at the external face of the inner membrane or in the outer membrane can be used to monitor the molecule on its journey to the cell surface³. First, the region of LPS that anchors it to the membrane, known as lipid A or endotoxin, along with its attached core sugars, is translocated to the external leaflet of the inner membrane. Here, a terminal O-sugar chain can be incorporated before the seven lipopolysaccharide transport proteins (LptA–G) take over (Fig. 1a). An ABC transporter complex (made up of LptF, LptG and two LptB proteins) in the inner membrane detaches LPS from the external leaflet and ushers it along a protein filament (made up of LptA and LptC), which tracks outward into the space between the two membranes⁴. Finally, a complex between LptD and LptE in the outer membrane functions as a

station that completes the track for delivering and inserting LPS into the external leaflet of the outer membrane.

Dong *et al.* and Qiao *et al.* present X-ray crystal structures of the complex formed between LptD and LptE from the bacteria *Salmonella typhimurium* and *Shigella flexneri*, respectively. The combined results of the two groups show that LptD is built primarily from structural segments known as β -strands that fold into

two domains — a β -jellyroll and a β -barrel. The smaller of the two, the β -jellyroll, extends away from the outer membrane and contains a greasy slot adapted to bind lipid A, while leaving the LPS sugar chain exposed. The same β -jellyroll fold is found in LptA and LptC, which interconnect to align their slots in an extended greasy groove that spirals across the space between the inner and outer membranes⁵. Thus, it seems that the LptD β -jellyroll is the terminal track for LPS delivery (Fig. 1b).

The structure of this LptD platform is positioned to deliver the LPS directly into the cavity of the much larger domain, the β -barrel, which is composed of 26 membrane-spanning β -strands, the greatest number of these strands ever described for a membrane β -barrel. The LptD β -barrel takes the shape of a kidney bean, with the lobe opposite the β -jellyroll plugged with another outer-membrane protein, LptE. This lipoprotein is positioned almost entirely inside the LptD β -barrel so as to interact most readily with the sugar chain of the delivered LPS — it has only its lipid-modified end anchored outside the β -barrel and within the inner leaflet of the outer membrane. The picture that emerges is one of LptE reorienting the LPS so that it stands vertically, with lipid A positioned transiently inside the open lobe of the β -barrel and the sugar chain protruding out to the cell surface. Although Dong and colleagues' structure had its β -jellyroll cleaved off and is

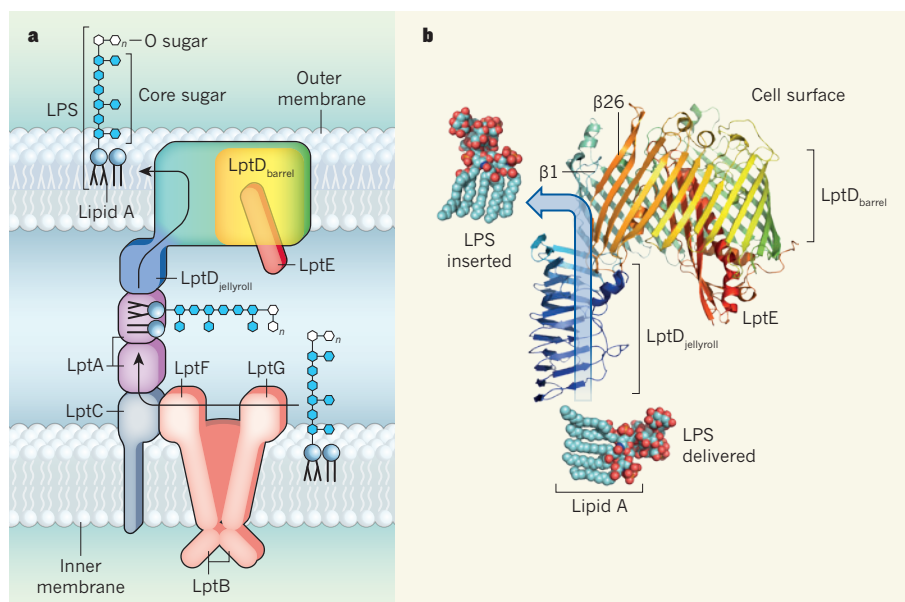


Figure 1 | Structure of the LptD–LptE protein complex. **a**, The bacterial outer-membrane component lipopolysaccharide (LPS) comprises a lipid A membrane anchor, a core sugar unit and a terminal O-sugar chain. The LPS is assembled in the cytoplasm and exported to the external leaflet of the inner membrane. Seven Lpt proteins (LptA–G) then transport it from the inner to the outer membrane. **b**, Dong *et al.*¹ and Qiao *et al.*² present the crystal structures of the complex formed between LptD and LptE proteins in the outer membrane of *Salmonella typhimurium* and *Shigella flexneri*, respectively, which reveal the route for LPS delivery and insertion into the outer membrane (blue arrow; the *S. flexneri* structure is shown). This structure shows that the sugars of LPS are delivered by the β -jellyroll region of LptD so as to engage first with LptE, which probably reorients the LPS into a vertical position. Insertion of the LPS into the outer-membrane external leaflet then occurs by lateral migration through an opening between the first (1) and final (26) β -strands in the LptD β -barrel.

of slightly lower resolution than that presented by Qiao *et al.*, it is accompanied by functional analyses that confirm an interesting prediction derived from earlier studies of two smaller β -barrel proteins, PagP (refs 6, 7) and FadL (refs 8, 9). The greasy membrane environment drives the formation of strong hydrogen bonds that 'zip' neighbouring β -barrel strands together, but the studies on PagP and FadL have revealed local gaps in β -strand hydrogen bonding that provide gateways for the lateral migration of lipids between their interiors and the external leaflet of the outer membrane.

Could a local gap in β -strand hydrogen bonding similarly support a lateral lipid-migration mechanism in LptD? The region of weakest β -strand hydrogen bonding in the LptD β -barrel is located nearest to the point of LPS passage, between the first and final β -strands (1 and 26, respectively). Dong *et al.* used molecular-dynamics simulations to show that these strands can separate. Furthermore, they observed that bacterial growth was blocked when they judiciously replaced amino acids in each of these β -strands to crosslink them together, supporting the idea that this gap in β -strand hydrogen bonding is necessary for optimal LPS transport. Qiao *et al.* similarly note that β -strands 1 and 2 are distorted in the *S. flexneri* LptD and that local dynamics exist both there and in β -strand 26. However, both structures show the gate in a closed state, and a structure of the complex with bound LPS might be needed to reveal further details of this lateral gating mechanism.

Although it is known that the ABC transporter drives LPS into the LptA–LptC filament¹⁰, it remains to be determined whether additional factors propel LPS out of the LptD–LptE complex so that it inserts into the outer membrane. LPS carries a strong net negative charge, and electrostatic repulsion between neighbouring molecules might help to move them along the filament; bridging of LPS by positively charged magnesium ions in the outer membrane might then facilitate its lateral egress from LptD.

Roughly 200 LptD–LptE complexes are evenly distributed over the surface of a typical bacterium, and each complex transports around 5 LPS molecules to the cell surface every second. A synthetic cyclic-peptide antibiotic is now in phase II clinical trials for treating infections with the bacterium *Pseudomonas aeruginosa* by targeting its LptD and blocking LPS transport^{11,12}. The structures of the LptD–LptE complexes from *S. typhimurium* and *S. flexneri* thus promise to facilitate the design of better antibiotics, in addition to further clarifying the mechanism by which LPS quite literally rolls out the barrel. ■

Russell E. Bishop is in the Department of Biochemistry and Biomedical Sciences, and at the Michael G. DeGroote Institute for Infectious Disease Research, McMaster University,

Hamilton, Ontario L8S 4K1, Canada.
e-mail: bishopr@mcmaster.ca

1. Dong, H. *et al.* *Nature* **511**, 52–56 (2014).
2. Qiao, S., Luo, Q., Zhao, Y., Zhang, X.-C. & Huang, Y. *Nature* **511**, 108–111 (2014).
3. Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. *Annu. Rev. Biochem.* **76**, 295–329 (2007).
4. Whitfield, C. & Trent, M. S. *Annu. Rev. Biochem.* **83**, 99–128 (2014).
5. Freinkman, E., Okuda, S., Ruiz, N. & Kahne, D. *Biochemistry* **19**, 4800–4806 (2012).
6. Khan, M. A. & Bishop, R. E. *Biochemistry* **48**,

9745–9756 (2009).

7. Cuesta-Seijo, J. A. *et al.* *Structure* **18**, 1210–1219 (2010).
8. Hearn, E. M., Patel, D. R., Lepore, B. W., Indic, M. & van den Berg, B. *Nature* **458**, 367–370 (2009).
9. Lepore, B. W. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 10121–10126 (2011).
10. Okuda, S., Freinkman, E. & Kahne, D. *Science* **338**, 1214–1217 (2012).
11. Srinivas, N. *et al.* *Science* **327**, 1010–1013 (2010).
12. Werneburg, M. *et al.* *ChemBioChem* **13**, 1767–1775 (2012).

This article was published online on 18 June 2014.

CLIMATE SCIENCE

Expulsion from history

Global warming is projected to force climatic variables in some places beyond the range of historical experience, perhaps permanently. A reassessment shows that this could begin sooner or much later than recently estimated.

SCOTT B. POWER

Earth will continue to warm under plausible 'business-as-usual' scenarios for future greenhouse-gas emissions¹, which assume that little, if anything, is done to abate emissions. This warming, in turn, will cause further widespread changes in the climate system¹. Future climatic conditions will also depend on naturally occurring climatic variability^{1,2}, a large component of which arises spontaneously within the climate system^{2,3}; the El Niño–Southern Oscillation is one example of such internal climate variability^{2,4}. In a paper published in *Nature* last year, Mora *et al.*⁵ estimated when permanent departures — or expulsions — from the historical range of variability will commence. Hawkins and

colleagues⁶ now dispute these estimates in a Brief Communications Arising published on *Nature's* website, and question the magnitude of the associated uncertainty provided by Mora and co-workers.

The timing of an expulsion (t_E) of a particular climate variable at a given location depends on the magnitude of the underlying warming, and on both the level and the timing of future natural variability (Fig. 1). However, the ability of climate scientists to predict internal variability beyond the next year or two is limited², and so the sequence of variability that will unfold during the remainder of the twenty-first century is largely unknown. This uncertain future sequence will either tend to bring expulsions forward in time, or delay their commencement.

Mora *et al.* and Hawkins *et al.* estimated t_E

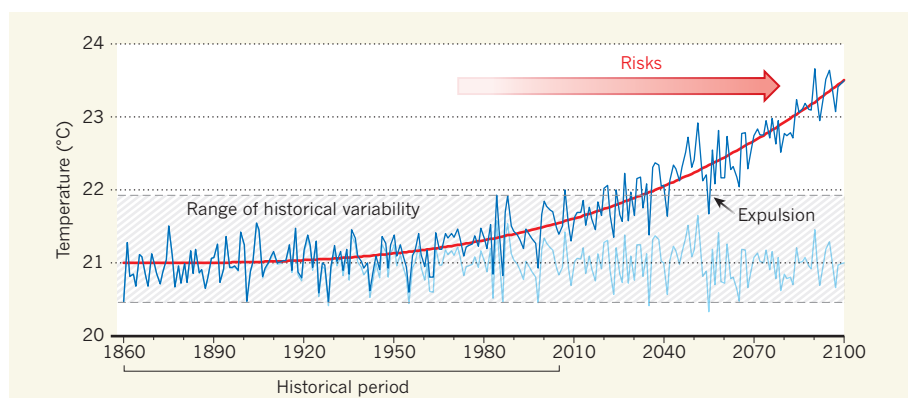


Figure 1 | Illustration of climatic expulsion. In this idealized example, temperature at a particular site (dark blue line, measured in °C) is plotted for the years 1860 to 2100. This graph represents the sum of natural temperature variations (light blue line) and of underlying warming associated with human activities (red line). Grey dashed lines indicate the range of variability over the historical period (1860–2005). The temperature remains permanently outside the historical range after time t_E (the expulsion time; black arrow). The red arrow indicates risks associated with climate change; the depth of shading represents the number and severity of risks, and correlates with the underlying warming. Mora *et al.*⁵ reported that expulsions for various climate variables can occur for many locations under 'business-as-usual' scenarios of greenhouse-gas emissions. Hawkins *et al.*⁶ now provide better estimates of t_E .

using information from many of the world's latest generation of climate models¹. The models generate a wide range of t_E values because each model has a different sequence of future natural variability, and differ in their sensitivity to the imposed increases in greenhouse-gas concentrations. Hawkins and colleagues contend that Mora and co-workers' estimates of t_E are, in some cases, too low (that is, some expulsions were estimated to occur too soon), and that the associated ranges of possible values are too narrow for many locations and climate variables.

Hawkins *et al.* argue that these problems arise for two reasons. First, Mora and colleagues analysed model results only up to the year 2100. In some places, expulsion seemed to occur before then, but in the remaining regions, Mora and colleagues set t_E to 2100. This approach underestimates t_E in regions where expulsion occurs after 2100, and misrepresents the situation in regions where expulsion might never happen. Second, Mora and co-workers based their discussion of the range of possible future values of t_E on a statistical measure of the uncertainty in the model-average value of t_E . Hawkins *et al.* regard this as an inappropriate measure of future uncertainty in t_E .

Mora *et al.* concede the first point in their response⁷ to Hawkins and colleagues' arguments (also published on *Nature's* website), but not the second. The source of this disagreement can be clarified with a simple analogy. Imagine that a doctor receives a text message from Harry, who is doing a school project on life in the nineteenth century: "What is the average age at death of people born in our city in 1860?" Having just read a report on the topic, the doctor promptly replies " 61 ± 1.2 years." This is the estimated average, together with an indication of the accuracy with which the average value is known.

A few days later, Harry sends a second message: "How old were they when they died?" The doctor, recalling her previous text, is about to re-type " 61 ± 1.2 years," but then realizes that this is a different question. The range of possibilities is substantially broader than ± 1.2 years. Some people died soon after birth, others lived beyond 80 years of age. She then replies: "0 to 80+ years."

In their original paper⁵, Mora *et al.* gave standard errors that provide (under certain assumptions⁸) confidence intervals associated with the model-average values of t_E . Although these are valid statistical measures, they are more closely related to the uncertainty in the doctor's first response than to that in the second. This is because the real world does not behave like the model-average — it will be much more like a single model realization with a particular sequence of future internal climate variability.

The problem is that we do not know which model sequence to choose, or even if any of the model sequences coincides with the one that will actually unfold. The best that we can do is to indicate the range of possible future

sequences. This is what Hawkins and colleagues do. Their ranges⁶ are much broader than those provided by Mora and co-workers⁵, and are analogous to the uncertainty in the doctor's second answer. Hawkins *et al.* conclude, for example, that there is more than an 85% chance that t_E at any location will fall outside the range provided by Mora and colleagues. In my opinion, the approach taken by Hawkins *et al.* provides a more appropriate estimate of the range of t_E that could occur in the real world over the coming century and beyond.

Despite these issues, important conclusions of Mora and co-workers' original paper remain valid. Expulsions are indeed expected to occur under business-as-usual scenarios over wide areas before 2100. These will tend to occur sooner under scenarios involving higher emissions, and are more likely to happen soonest in regions that include biodiversity hot-spots and many low-income countries^{3,9}.

Highlighting the genuine risk of expulsions⁵ and providing better estimates of when they can occur⁶ are both valuable. It is crucial to realize, however, that more-modest changes occurring well before an expulsion can also be a major concern. This is because such changes can still be large and rapid enough¹⁰ to have severe impacts on humans and ecosystems¹¹. ■

Scott B. Power is at the Bureau of Meteorology, GPO Box 1289, Melbourne, Victoria 3001, Australia.

IMMUNOLOGY

Fixing the odds against tuberculosis

A study that defines the interactions between three immunity-regulating molecules — type 1 interferon, interleukin-1 and prostaglandin E_2 — will help to explain the variable outcomes of tuberculosis infections. [SEE LETTER P.99](#)

SAMUEL M. BEHAR
& CHRISTOPHER M. SASSETTI

Billions of people worldwide are infected with *Mycobacterium tuberculosis*. However, only 5–10% of these individuals will develop clinically evident disease — a likelihood similar to winning a game of roulette by betting on a single number. Many genetic and environmental variables are known to increase the risk of progression to clinical tuberculosis (TB). Understanding how these factors interact to determine the outcome of infection could facilitate the design of better therapies and the targeting of their use to those at highest risk. On page 99 of this issue, Mayer-Barber *et al.*¹ describe an immune network that may have a

e-mail: s.power@bom.gov.au

1. Intergovernmental Panel on Climate Change. 'Summary for Policymakers' in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 3–29 (Cambridge Univ. Press, 2013).
2. Kirtman, B. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 11, 953–1028 (Cambridge Univ. Press, 2013).
3. Lorenz, E. N. *J. Atmos. Sci.* **20**, 130–141 (1963).
4. Christensen, J. H. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 14, 1217–1308 (Cambridge Univ. Press, 2013).
5. Mora, C. *et al.* *Nature* **502**, 183–187 (2013).
6. Hawkins, E. *et al.* *Nature* **511**, E3–E4 (2014).
7. Mora, C. *et al.* *Nature* **511**, E5–E6 (2014).
8. Power, S. B., Delage, F., Colman, R. & Moise, A. *J. Clim.* **25**, 3792–3809 (2012).
9. Mahlstein, I., Knutti, R., Solomon, S. & Portmann, R. W. *Environ. Res. Lett.* **6**, 034009 (2011).
10. Settele, J. *et al.* in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. *et al.*) Ch. 4 (Cambridge Univ. Press, 2014).
11. Intergovernmental Panel on Climate Change. 'Summary for Policymakers' in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C. B. *et al.*) 1–32 (Cambridge Univ. Press, 2014).

central role in determining TB outcomes and could be manipulated to tip the odds in favour of host resistance.

Numerous mediators of immunity have been implicated in host resistance to TB. These regulators of the immune response come in many forms, including secreted proteins, such as cytokines, and lipid-derived signalling molecules, such as eicosanoids. Most of these were discovered when their complete removal or marked overproduction altered the outcome of infection. Such studies^{2–4} identified roles for the cytokines interleukin-1 (IL-1) and type I interferon (IFN) and the eicosanoid prostaglandin E_2 (PGE₂) in the pathogenesis of TB.

Mayer-Barber *et al.* take the step of placing

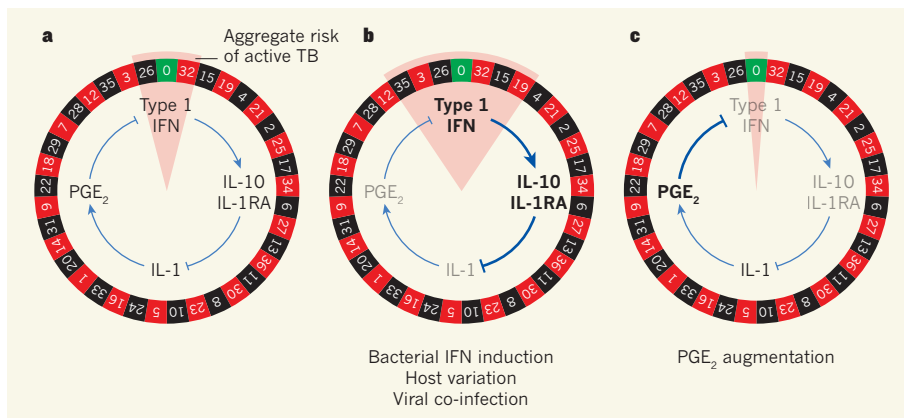


Figure 1 | Factors influencing the outcome of tuberculosis infection. Mayer-Barber *et al.*¹ provide evidence from animal models and human clinical data that the depicted network of immune-regulatory molecules influences the gamble that is infection with *Mycobacterium tuberculosis*. **a**, Type I interferon (IFN) promotes disease by inducing interleukin-10 (IL-10) and IL-1 receptor antagonist (IL-1RA). These mediators inhibit IL-1-dependent prostaglandin E₂ (PGE₂), which would otherwise restrict the intracellular growth of the bacteria. **b**, Although factors that increase the risk of developing tuberculosis probably act by diverse mechanisms, many could influence the activity or balance of this immune circuit, particularly by accentuating IFN production. **c**, By contrast, Mayer-Barber *et al.* show that PGE₂ supplementation can reduce disease severity in infected mice that have high IFN levels.

these crucial players into their regulatory context, and discover that they form a cross-regulatory network. In *M. tuberculosis*-infected mice, the authors find that IL-1 induces PGE₂, which enhances the antimicrobial activity of immune cells called macrophages. IFN inhibits the production and activity of IL-1 and is itself inhibited by PGE₂ (Fig. 1). This pathway provides a mechanistic framework to explain the previous observations that IFN exacerbates TB⁵, whereas IL-1 and PGE₂ are generally protective^{2,4}. The mechanisms by which IL-1 and PGE₂ inhibit bacterial growth remain to be defined, but are probably related, at least in part, to the inhibition of macrophage necrotic cell death⁶.

The authors also observed that alterations in this IL-1–PGE₂–IFN network were associated with more severe disease in humans, providing validation for their animal studies. These clinical data significantly further our ability to delineate distinct disease states using immunological markers. Previous studies demonstrated an association between IFN responses and active TB in humans⁷. Adding other members of this immune network to the analysis has refined the resolution of this approach, allowing Mayer-Barber and colleagues to differentiate patients with mild disease from those classified as asymptomatic. This type of rationally designed biomarker panel may prove useful in the design of therapeutic trials.

The proposed IL-1–PGE₂–IFN pathway could be interpreted as a feedback loop that maintains homeostasis in the body, because excess activity of each mediator triggers an inhibitory response, returning the system to balance. However, the situation is probably more complicated during chronic TB. For example, the kinetics of each mediator's expression may be crucial. IFN production tends to be rapid and transient, whereas IL-1 is typically associated with chronic inflammatory

states. Indeed, although IL-1 has a protective role during early infection, its persistent production causes disease-associated symptoms⁸.

In addition, IFN expression is remarkably stochastic at the single-cell level⁹, suggesting that IFN-expressing cells could be unevenly distributed in the lung tissue. The observation that different TB lesions in the same lung evolve independently¹⁰ supports the idea that the spatial distribution of these mediators is functionally variable. Perhaps most significantly, these innate immune responses influence and are influenced by the T-cell-based adaptive immune response, which can vary throughout a prolonged chronic infection. Thus, although IL-1, PGE₂ and IFN can be depicted together as a simple, self-regulating pathway, any number of factors could tip the balance away from protective PGE₂ and towards pathological IFN production.

Several genetic and environmental variables are known to increase the odds of developing TB, and might act by modulating the IL-1–PGE₂–IFN network. Pre-existing genetic variation in both the host and the pathogen probably have significant roles. For example, highly pathogenic strains of *M. tuberculosis* are known to express an IFN-inducing phenolic glycolipid, and infection with these strains causes increased IFN production¹¹. Similarly, human genetic variations in eicosanoid biosynthesis are associated with TB¹², and variations in IL-1 and IFN production have been found to influence a variety of other infectious diseases. Even non-genetic factors such as viral co-infection could promote TB pathogenesis through IFN induction¹³.

Are we destined to remain at the mercy of this game of chance? An exciting aspect of the proposed IL-1–PGE₂–IFN network is the implied strategy for intervention. In a mouse model of severe TB that is driven by high IFN

levels, Mayer-Barber *et al.* found that PGE₂ augmentation suppressed IFN, ameliorated disease and promoted bacterial killing. This is an outstanding example of fundamental biological insight leading to a new therapeutic opportunity. Of course, the potential for this treatment strategy depends on the demonstration that IFN is a key driver of human TB, which is currently unclear. In addition, it is difficult to imagine that a host-directed therapy such as this would be administered without simultaneous antimicrobial chemotherapy, and the benefit of PGE₂ in this context would need to be assessed. Nevertheless, when we consider that more than a million people die of TB each year and the continual emergence of antibiotic-resistant *M. tuberculosis* strains that are insensitive to standard therapy, any chance to fix the game in favour of the host should be considered. ■

Samuel M. Behar and Christopher

M. Sasseti are in the Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. C.M.S. is also at the Howard Hughes Medical Institute, Chevy Chase, Maryland. e-mails: samuel.behar@umassmed.edu; christopher.sasseti@umassmed.edu

1. Mayer-Barber, K. D. *et al.* *Nature* **511**, 99–103 (2014).
2. Mayer-Barber, K. D. *et al.* *J. Immunol.* **184**, 3326–3330 (2010).
3. Ordway, D. *et al.* *J. Immunol.* **179**, 522–531 (2007).
4. Divangahi, M., Desjardins, D., Nunes-Alves, C., Remold, H. G. & Behar, S. M. *Nature Immunol.* **11**, 751–758 (2010).
5. Antonelli, L. R. V. *et al.* *J. Clin. Invest.* **120**, 1674–1682 (2010).
6. Divangahi, M. *et al.* *Nature Immunol.* **10**, 899–906 (2009).
7. Berry, M. P. R. *et al.* *Nature* **466**, 973–977 (2010).
8. Mishra, B. B. *et al.* *Nature Immunol.* **14**, 52–60 (2013).
9. Zhao, M., Zhang, J., Phatnani, H., Scheu, S. & Maniatis, T. *PLoS Biol.* **10**, e1001249 (2012).
10. Lin, P. L. *et al.* *Nature Med.* **20**, 75–79 (2014).
11. Manca, C. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 5752–5757 (2001).
12. Tobin, D. M. *et al.* *Cell* **140**, 717–730 (2010).
13. Redford, P. S. *et al.* *J. Infect. Dis.* **209**, 270–274 (2014).

This article was published online on 25 June 2014.

RETRACTION

In view of the fact that the authors of 'Stimulus-triggered fate conversion of somatic cells into pluripotency' (H. Obokata *et al.* *Nature* **505**, 641–647; 2014) and 'Bidirectional developmental potential in reprogrammed cells with acquired pluripotency' (H. Obokata *et al.* *Nature* **505**, 676–680; 2014) are retracting their reports (see page 112), I wish to retract the News & Views article, 'Potency unchained' (*Nature* **505**, 622–623; 2014), which dealt with these studies and was based on the accuracy and reproducibility of their data. **Austin Smith**

Patterning and post-patterning modes of evolutionary digit loss in mammals

Kimberly L. Cooper^{1†*}, Karen E. Sears^{2*}, Aysu Uygur^{1*}, Jennifer Maier², Karl-Stephan Baczkowski³, Margaret Brosnahan⁴, Doug Antczak⁴, Julian A. Skidmore⁵ & Clifford J. Tabin¹

A reduction in the number of digits has evolved many times in tetrapods, particularly in cursorial mammals that travel over deserts and plains, yet the underlying developmental mechanisms have remained elusive. Here we show that digit loss can occur both during early limb patterning and at later post-patterning stages of chondrogenesis. In the ‘odd-toed’ jerboa (*Dipus sagitta*) and horse and the ‘even-toed’ camel, extensive cell death sculpts the tissue around the remaining toes. In contrast, digit loss in the pig is orchestrated by earlier limb patterning mechanisms including downregulation of *Ptch1* expression but no increase in cell death. Together these data demonstrate remarkable plasticity in the mechanisms of vertebrate limb evolution and shed light on the complexity of morphological convergence, particularly within the artiodactyl lineage.

Tetrapod limbs evolved adaptations for running, swimming, flying and many other tasks, each reflected in functional modifications to their morphology. Digit reduction, a decrease in the number of digits from the basal pentadactyl (five-digit) morphology, arose repeatedly in tetrapod evolution¹. There are two plausible developmental mechanisms by which this could take place. The first would be to specify fewer digit primordia during the time when developmental fates are patterned in the early limb bud. The second would be to initially organize the limb bud in a normal pentadactyl pattern but then fail to elaborate the full set of digits by resculpting the nascent limb through differential proliferation or cell death.

To date, the molecular developmental mechanism of evolutionary digit reduction has been explored in only one tetrapod group, the skinks of the genus *Hemiergis*. Distinct species of *Hemiergis* range in digit number from two to five^{2,3}, with evolutionary progression to fewer digits correlating with increasingly early termination of Sonic hedgehog (*Shh*) expression in the posterior limb bud⁴. *Shh* serves a dual purpose in limb development, both to pattern the digits and to expand the hand and foot plates to allow for the formation of a full complement of digits^{5–7}. Experimental truncation of the developmental timing of *Shh* expression removes digits in the reverse order of their formation⁷, consistent with a mechanism first suggested by Alberch and Gale⁸, thus providing a convenient way to evolutionarily tweak digit number without disturbing the overall structure of the limb. However, this mechanism would not, in a simple manner, generate the symmetrical reduction of anterior (pre-axial) and posterior (post-axial) digits seen, for example, in the evolution of the horse lineage.

To investigate how digit reduction evolved in other adaptive contexts we examined the mode of digit loss in a bipedal three-toed rodent and in three ungulates: the single-toed horse, an odd-toed ungulate or perissodactyl, and the pig with four toes and the camel with two, each representing the even-toed ungulates or artiodactyls (Fig. 1a, b).

Mechanisms of digit loss in the three-toed jerboa

We first focused on the three-toed jerboa, *D. sagitta* (Fig. 1f). This species has several advantages in identifying meaningful alterations to ancestral developmental mechanisms. First, it has a close evolutionary relationship to the laboratory mouse and to a five-toed species of jerboa, *Allactaga elater*⁹ (Fig. 1d). Moreover, digit loss in *D. sagitta* is limited to

the hindlimb whereas forelimbs maintained five fully formed fingers^{10,11}. This provides a unique opportunity to identify differences specific to morphological divergence of the hindlimb among a potential abundance of species-specific modifications shared in the development of both sets of paired appendages.

In the adult *D. sagitta*, the three central metatarsals are fused into a single element that trifurcates distally and articulates with each of the three digits¹⁰. However, in the neonate, alcian blue and alizarin red staining of the chondrogenic skeleton reveals that the three complete digits and their associated metatarsals are flanked by small, truncated cartilage remnants of the first and fifth metatarsals (Fig. 1c and Extended Data Fig. 1). This suggests that at least the proximal-most portion of each of the five digit rays is patterned early in development and that digits I and V are either not fully patterned distally or are truncated at a subsequent differentiation stage.

To gain a better sense of when the patterning and/or morphogenesis of the lateral digits begins to diverge in the three-toed jerboa hindlimb, we compared the contours of various staged limb buds between mice and *D. sagitta*. We found that when scaled for size, the forelimbs of mice and three-toed jerboas are consistently identical in morphology. In contrast, the *D. sagitta* hindlimb starts to be noticeably narrower as early as 11.5 days post conception (dpc) (Extended Data Fig. 2).

Accordingly, we conducted an expression screen of a series of genes known to be involved in limb patterning just before and at the time of morphological divergence in hindlimb bud shape. None of the patterning genes we examined showed an obvious difference in expression in the *D. sagitta* hindlimb, including *Shh*, *Ptch1*, *Gli1* and *HoxD13* (Fig. 2a, b). Turning to post-patterning stages, cell proliferation was assessed by phospho-histone H3 antigen detection. However, we did not see a decrease in proliferation in the hindlimb of the three-toed jerboa, either at early stages of autopod expansion or later during digit outgrowth in any domain of the developing limbs (Extended Data Fig. 3). In contrast, we saw derived expanded domains of TdT-mediated dUTP nick end labelling (TUNEL)-positive nuclei, a marker for programmed cell death, specific to the jerboa hindlimb as early as 12.5 dpc (Extended Data Fig. 6). These domains further expand by 13.5 dpc to encompass all of the tissue distal to what would become the truncated cartilage condensations (Fig. 3b). Thus

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Animal Biology, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA. ³École Normale Supérieure de Lyon, 69007 Lyon, France. ⁴Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA. ⁵The Camel Reproduction Centre, Dubai, United Arab Emirates. [†]Present address: Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

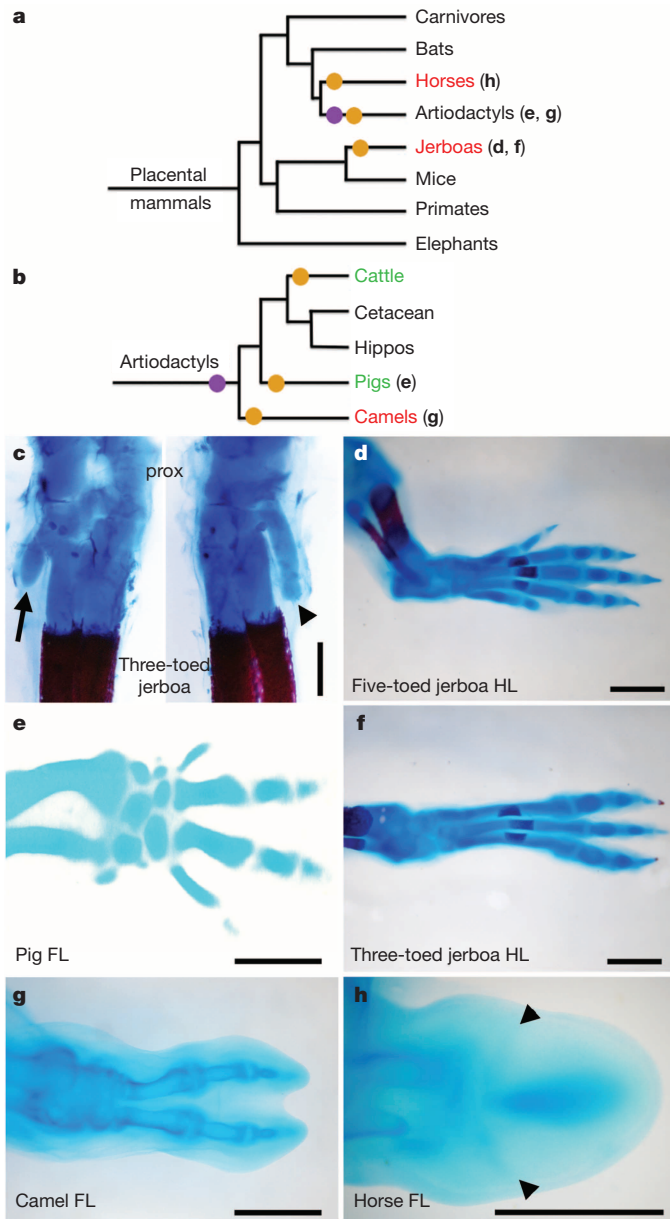


Figure 1 | Convergent evolution of the embryonic limb skeleton in multiple mammalian species. **a, b,** Phylogeny of mammals (**a**) and of artiodactyls (**b**) representing the major groups that have independently lost digits, based on ref. 37. Parenthetical lettering references skeletons in accompanying panels. Orange circles indicate an evolutionary incidence of digit loss. Purple circles represent the shift from mesaxonic to paraxonic limbs in basal artiodactyls. Species that sculpt the limb by cell death are highlighted in red, and those that show a restriction of *Ptch1* expression are highlighted in green. **c,** Alcian blue and alizarin red stained skeleton of postnatal day 0 three-toed jerboa, *D. sagitta*, with the ankle (proximal) at the top. Posterior view (left) highlights the truncated fifth metatarsal (arrow). Anterior view (right) highlights the truncated first metatarsal (arrowhead) ($n = 4$ embryos). **d,** Alcian blue stained skeletons of the approximately 16 dpc five-toed jerboa, *A. elater*, hindlimb (HL) ($n = 3$ embryos). **e,** 30 dpc pig forelimb (FL) ($n = 2$ embryos). **f,** Approximately 16 dpc *D. sagitta* hindlimb ($n = 2$ embryos). **g,** 50 dpc camel hindlimb ($n = 1$ embryos). **h,** 34 dpc horse forelimb ($n = 1$ embryo). Scale bar, 0.5 mm. (**c, d, f**); scale bar, 1 mm (**e, g, h**).

digit loss in this species appears to result from the sculpting of anterior (pre-axial) and posterior (post-axial) tissues at the distal ends of properly patterned nascent digits.

Apoptosis is used in basal tetrapods to sculpt the digits, removing interdigital tissue late in limb development¹². This suggests that a

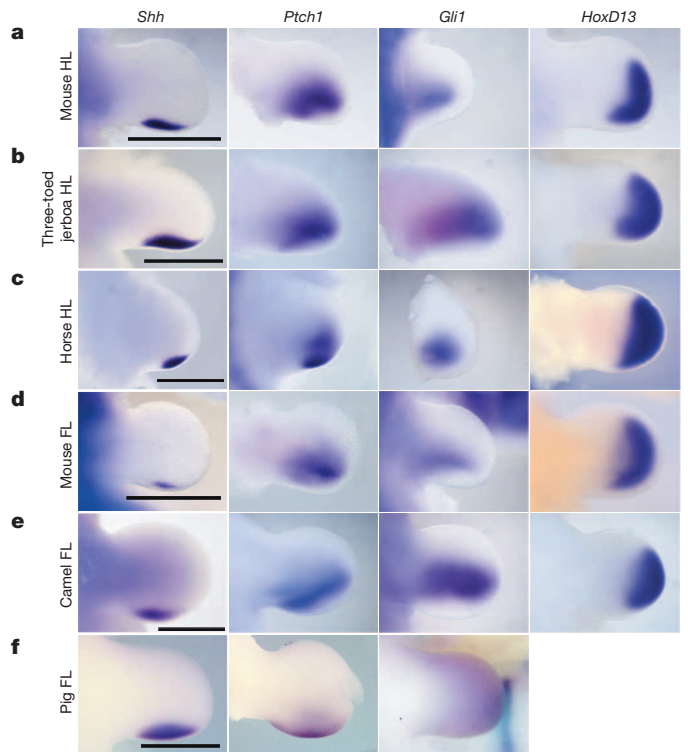


Figure 2 | Expression of early patterning genes: *Shh*, *Ptch1*, *Gli1* and *HoxD13*. **a,** Mouse hindlimb ($n = 3$ embryos per gene). **b,** Three-toed jerboa, *D. sagitta*, hindlimb ($n = 3$ embryos per gene). **c,** Horse hindlimb ($n = 1$ embryo for *Shh* and $n = 2$ embryos for *Ptch1*, *Gli1*, and *HoxD13*). **d,** Mouse forelimb ($n = 3$ embryos per gene). **e,** Camel forelimb ($n = 1$ each for *Shh* and *HoxD13* and $n = 2$ embryos for *Ptch1* and *Gli1*). **f,** Pig forelimb. Scale bar, 1 mm.

potential evolutionary route for achieving cell death in the *D. sagitta* hindlimb digit I and V primordia might be through co-option of the apoptotic pathways normally used to direct interdigital cell death. The transcription factor *Msx2* is strongly expressed in the interdigital tissue of the embryonic mouse and chicken¹³, and retroviral misexpression in chicken embryos induces a striking increase in cell death and loss of cartilaginous digit condensations^{14,15}. We found that *Msx2* is strongly expressed in the *D. sagitta* hindlimb in tissue surrounding and distal to the truncated first and fifth metatarsals and completely overlaps with domains of TUNEL-positive nuclei (Fig. 4a–c). In different contexts within the limb bud, the secreted protein *Bmp4* can act both upstream and downstream of *Msx2*^{15,16}. We observe a transient spatial increase of *Bmp4* expression specific to the *D. sagitta* hindlimb autopod starting at 12 dpc that resolves at 12.5 dpc into two strong and discrete domains of expression precisely prefiguring the proximal positions of the first and fifth digits (Extended Data Fig. 4). However, *Msx2* is expanded in the *D. sagitta* hindlimb before expanded *Bmp4* expression, as early as 11 dpc (Extended Data Fig. 5). This is when the *D. sagitta* hindlimb first shows signs of narrowing relative to limbs that will develop five digits (Extended Data Fig. 2), consistent with altered *Msx2* expression potentially being the primary causal mechanism of digit loss in this species.

As the interdigital tissue begins to undergo apoptosis during mouse limb development, *Fgf8* expression is lost in the overlying apical ectodermal ridge (AER), whereas *Fgf8* expression is maintained above the growing digits (Fig. 5a). *Fgf8* is both necessary and sufficient for digit outgrowth in mouse and chicken embryos^{17–20}. From about 12.75 dpc in the *D. sagitta* hindlimb, *Fgf8* expression regresses away from the posterior and then anterior AER as well as the interdigital domains, persisting only over the digits that continue to develop to completion (Fig. 5a, b and Extended Data Fig. 6).

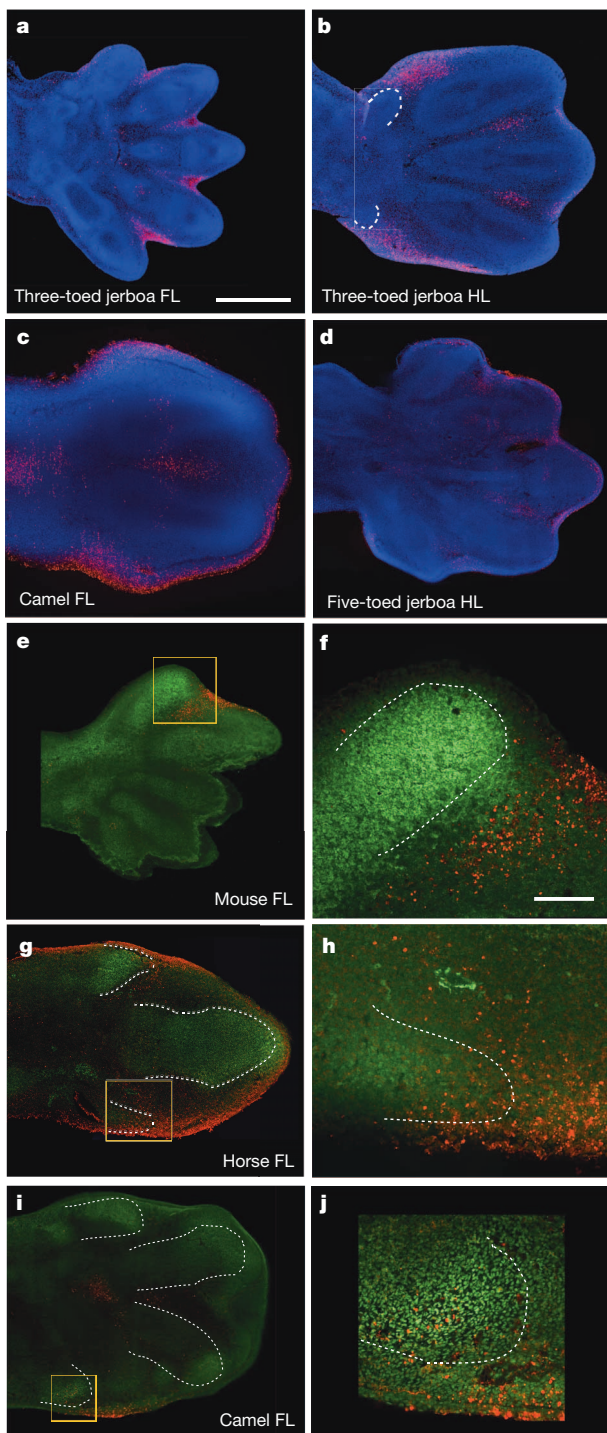


Figure 3 | Patterns of cell death in each mammalian limb. Fluorescent detection of DAPI (4',6-diamidino-2-phenylindole) is blue, Sox9 immunohistochemistry is green, and TUNEL is red. **a**, Approximately 13.5 dpc three-toed jerboa, *D. sagitta*, forelimb. **b**, Approximately 13.5 dpc *D. sagitta* hindlimb (white dashed line indicates truncated metatarsals I and V) ($n = 2$ embryos). **c**, 45 dpc camel forelimb ($n = 1$ embryo). **d**, Approximately 13.5 dpc five-toed jerboa, *A. elater*, hindlimb ($n = 2$ embryos). **e**, Mouse at 13.5 dpc with Sox9 and TUNEL ($n = 3$ embryos). **f**, Magnification of boxed region in (**e**). **g**, 34 dpc horse forelimb ($n = 3$ embryos). **h**, Magnification of boxed region in (**g**). **i**, 42 dpc camel forelimb ($n = 2$ embryos). **j**, Magnification of boxed region in (**i**). Scale bar, 0.5 mm in **a** is also the same for **b–d**, **e**, **g** and **i**. Scale bar, 0.1 mm in **f** is also the same for **h** and **j**.

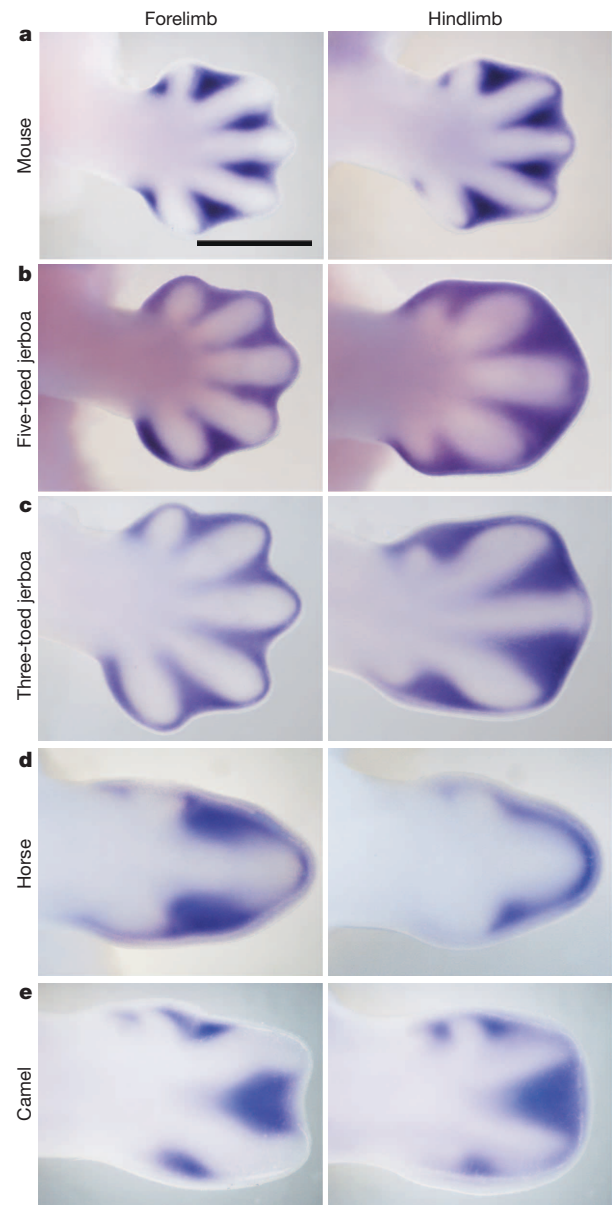


Figure 4 | Expression of *Msx2* at the start of digit chondrogenesis. **a–e**, Forelimb and hindlimb of 13 dpc mouse ($n = 3$ embryos) (**a**); approximately 13 dpc five-toed jerboa, *A. elater* ($n = 2$ embryos) (**b**); approximately 13 dpc three-toed jerboa, *D. sagitta* ($n = 2$ embryos) (**c**); 34 dpc horse ($n = 2$ embryos) (**d**); 42 dpc camel ($n = 1$ embryo) (**e**). Scale bar, 1 mm.

Convergence of post-pattern sculpting in the horse

The three-toed jerboa hindlimb remarkably resembles the limb structure of some of the early ancestral equine species with three toes²¹. To test possible mechanisms for digit reduction in the horse, we once again started by examining the expression of genes known to be involved in patterning the early limb bud. We observed no obvious differences in expression of *Shh*, *Ptch1*, *Gli1* or *HoxD13* relative to those previously described in mice (Fig. 2c). In contrast, we did observe TUNEL-positive nuclei entirely surrounding the central toe and within the distal ends of nascent Sox9⁺ truncated condensations of metacarpals 2 and 4 (Fig. 3g, h), a condition not observed in mouse (Fig. 3e, f). Moreover, we found expanded *Msx2* expression in domains correlating with those regions of anterior and posterior cell death (Fig. 4d). We also observed increased posterior expression of *Msx2* earlier in development (Extended Data Fig. 5) and distal expansion of *Bmp4* in both forelimbs and hindlimbs (Extended Data Fig. 4) similar to *D. sagitta* hindlimbs. *Fgf8* expression is

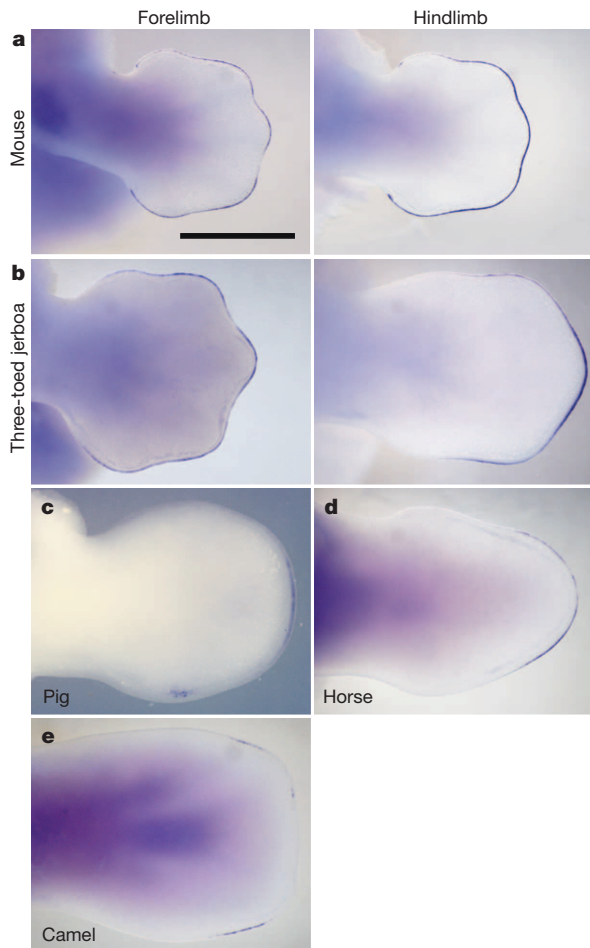


Figure 5 | *Fgf8* expression is restricted to the AER overlying nascent digits. **a–e**, Forelimb and hindlimb of 13 dpc mouse ($n = 3$ embryos) (**a**); approximately 13 dpc three-toed jerboa, *D. sagitta* ($n = 3$ embryos) (**b**); 25 dpc pig ($n = 2$ embryos) (**c**); 34 dpc horse hindlimb ($n = 2$ embryos) (**d**); 42 dpc camel forelimb ($n = 1$ embryo) (**e**). Scale bar, 1 mm.

also maintained in the horse AER only over the nascent central digit III (Fig. 5d). Thus, in the horse as in the three-toed jerboa, digit reduction appears to have a post-patterning contribution involving expanded domains of lateral apoptosis, possibly in part through similar molecular mechanisms. It is probable that mechanisms yet to be identified eliminate the first and fifth digits while a jerboa-like carving away of digits II and IV occurs by transforming cells from a chondrogenic to an apoptotic fate. A more extensive investigation of early patterning may be worthwhile with additional precisely staged early horse embryos.

Plasticity in artiodactyl digit loss

The even-toed ungulates present yet another opportunity to explore the possible convergence of digit reduction mechanisms in the context of additional skeletal remodelling. The distal artiodactyl limb has shifted the central axis of symmetry from digit III in the ancestral mesaxonic limb to a derived paraxonic limb where the axis of symmetry runs through the interdigital space between digits III and IV²². To explore whether digit loss in these species occurs via patterning and/or post-patterning changes, we obtained embryos from two species of artiodactyls, the pig and camel. While this work was in progress, we learned of similar studies by Lopez-Rios *et al.*²³ in a third artiodactyl species with convergent digit loss to two toes, the cow. The accompanying paper²³ identifies a gene regulatory control region for *Ptch1* expression in the limb that is altered in the cow. The resulting expression of *Ptch1* is reduced and more posteriorly restricted than in non-artiodactyl species. One role that *Ptch1* expression serves is to restrict the movement of the morphogen *Shh* across the limb bud^{24,25}. As a

consequence of the change in *Ptch1* expression in the cow, *Shh* targets, including *Gli1* and the *Hoxd* genes, are expressed more uniformly across the limb bud²³. Mice in which *Ptch1* expression is reduced in the limb display similar changes in downstream genes and a concomitant shift in the central axis of the limb to the space between digits III and IV and loss of the first digit²⁶. Importantly, after learning of our results with the three-toed jerboa and horse, Lopez-Rios and colleagues looked closely and saw no evidence of expanded apoptosis in the developing cow limb²³. Together these results suggest that, as in *Hemiergis*, the even-toed ungulates might have lost digits through a *Shh*-dependent patterning mechanism, albeit by a different genetic alteration, allowing the digits to be lost in an asymmetrical manner in the artiodactyls.

As would be expected if mutations affecting *Ptch1* regulation play a prominent role in artiodactyl limb evolution, we find *Ptch1* expression in the pig is also posteriorly restricted and downregulated concomitant with an upregulation of *Gli1* (Fig. 2f). Furthermore, as in the cow, there is no evidence of increased cell death in developing pig limbs²⁷. Surprisingly, however, *Ptch1* expression is not downregulated and restricted in the camel and is instead expressed much like non-artiodactyls (Fig. 2e). In addition, *Shh*, *Gli1*, and *HoxD13* exhibit ancestral patterns of expression indicating early patterning of the digit field by this subset of molecules is conserved in the camel (Fig. 2e). In contrast, when we examined patterns of cell death in the camel, we found expansive apoptosis throughout outgrowths of tissue flanking digits III and IV at 45 dpc (Fig. 3c) as well as at 42 dpc within small Sox9⁺ pre-cartilaginous nodules in the positions of missing digits II and V (Fig. 3i, j). As in the three-toed jerboa and horse, this correlates with domains of *Msx2* expression in the anterior and posterior limb bud at the time of digit condensation (Fig. 4e), though earlier expression of *Bmp4* and *Msx2* does not correlate, suggesting a distinct initiating mechanism for camel (Extended Data Figs 4 and 5).

Regardless of the mechanism by which digit loss occurs, at patterning or post-patterning stages, *Fgf8* expression is lost from the AER anterior and posterior to the digits that continue to develop in the pig, camel (Fig. 5c, e), and cow²³, as seen with the three-toed jerboa and horse. Loss of *Fgf8* from the anterior and posterior AER in the pig and cow, two species that lack expanded domains of cell death, uncouples this expression change from the direct cause of apoptosis and may instead reflect an independent requirement for its elimination to allow for digit termination in all species.

Discussion

These data indicate that at least two mechanisms of digit reduction are employed in the even-toed ungulates, one (exemplified by the pig and cow) involving changes in early patterning by *Shh* and not involving apoptosis, and a distinct mechanism (seen in camels) involving changes in domains of apoptosis that resculpt the limb after the patterning phase. These data do, however, present a paradox in the context of the well-established artiodactyl phylogeny and fossil record (Fig. 1b). Although the morphology of the cow and pig is remarkably similar to the mouse phenotype when *Ptch1* is lost from the limbs, both in the reduction of digits and shift in the symmetry of toes to the interdigit of III–IV²³, it can not have been responsible for both phenotypes in the artiodactyls as they occurred at different stages evolutionarily. The change in *Ptch1* regulation seen in both pigs and cattle indicates that it was probably present in their last common ancestor. As such, it cannot have been solely responsible for the loss of digit 1, as this occurred convergently in these two lineages. Indeed, digit reduction occurred at multiple independent times within the artiodactyl clade (Fig. 1b, orange circles), as the stem group of each major lineage was pentadactyl at least in the forelimb²⁸. The common ancestor of pigs and cattle would also have been ancestral to the hippos and their cetacean relatives, the dolphins and whales (Fig. 1b). Like extinct basal artiodactyls, hippos and basal cetaceans have a relatively small first digit^{22,27,29}. Thus, a restriction of *Ptch1* in a basal member of the group including pigs, hippos, cetaceans, and cattle may have served to reduce the size of the first digit and predispose the limb to further digit loss.

Perhaps even more striking is the absence of altered *Ptch1* regulation in the camel. Without this information, one might have speculated that the

Ptch1 mutation was responsible for the reorientation of the axis of symmetry in artiodactyls similar to the mutant mouse. However, the shift in the position of digits from mesaxonic to paraxonic is believed to be ancestral to the split of modern artiodactyl suborders (and indeed is a defining characteristic trait for this clade^{30–32}) (Fig. 1b, purple circle). Given the camel evidence, one has to either conclude that the shift actually arose independently in the ancestors of the camels and those of other artiodactyl lineages, or alternatively, any role *Ptch1* may have in the establishment of digit position in the pig and cow arose secondary to a separate mechanism established before the split of camels and their relatives.

The identification of several distinct molecular and cellular mechanisms of digit loss with recurring motifs suggests the developmental program of the tetrapod limb is fairly plastic. This would have provided some flexibility to allow adaptation in different circumstances and ultimately contributed to the diversity of limbs seen today.

METHODS SUMMARY

All embryos were collected in accordance with the appropriate Institutional Animal Care and Use Committee guidelines. Skeletal stainings were performed as previously described^{33,34}. Whole mount *in situ* hybridizations were performed for mouse, three and five-toed jerboa, horse, and camel as in ref. 35 and for pig embryos as in ref. 36. Riboprobes were generated by PCR amplifying from cDNA of the appropriate species, cloning into pGEM-T-Easy (Promega), sequence verification, and expression testing first in mouse embryos. Primers used for probe generation and accession numbers are listed in Extended Data Table 1. For TUNEL and immunohistochemistry (IHC), embryos were embedded in paraffin or gelatin and sectioned at 12 or 80 µm thickness. TUNEL was performed using the TMRd *In situ* Cell Death Detection Kit (Roche). Sox9 and PH3 IHC were each performed after boiling antigen retrieval in citrate buffer using a 1:500 dilution of rabbit anti-Sox9 (Millipore AB5535) or 1:200 dilution of rabbit anti PH3 (Cell Signal #9701) followed by a Cy2 or Alexa594 conjugated secondary respectively. Each experiment was performed in at least two limb buds (1–2 embryos per experiment). Fluorescent images were captured by confocal microscopy, and images of sequential sections were overlaid in NIH ImageJ.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 December 2013; accepted 22 May 2014.

Published online 18 June 2014.

- Lande, R. Evolutionary mechanisms of limb loss in tetrapods. *Evolution* **32**, 73–92 (1978).
- Skinner, A., Lee, M. S. & Hutchinson, M. N. Rapid and repeated limb loss in a clade of scincid lizards. *BMC Evol. Biol.* **8**, 310 (2008).
- Shapiro, M. D. Developmental morphology of limb reduction in *Hemiergis* (squamata: scincidae): chondrogenesis, osteogenesis, and heterochrony. *J. Morphol.* **254**, 211–231 (2002).
- Shapiro, M. D., Hanken, J. & Rosenthal, N. Developmental basis of evolutionary digit loss in the Australian lizard *Hemiergis*. *J. Exp. Zool. B Mol. Dev. Evol.* **297B**, 48–56 (2003).
- Harfe, B. D. et al. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
- Towers, M., Mahood, R., Yin, Y. & Tickle, C. Integration of growth and specification in chick wing digit-patterning. *Nature* **452**, 882–886 (2008).
- Zhu, J. et al. Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell* **14**, 624–632 (2008).
- Alberch, P. & Gale, E. A. Size dependence during the development of the amphibian foot. Colchicine-induced digital loss and reduction. *J. Embryol. Exp. Morphol.* **76**, 177–197 (1983).
- Walker, E. P. *Mammals of the World* (John Hopkins Press, 1964).
- Shenbrot, G. I., Sokolov, V. E. & Heptner, V. G. *Jerboas: Mammals of Russia and Adjacent Regions* (Science Publishers, 2008).
- Cooper, K. L. The lesser Egyptian jerboa, *Jaculus jaculus*: a unique rodent model for evolution and development. *Cold Spring Harb. Protoc.* **2011**, pdb.emo066704 (2011).
- Zuzarte-Luis, V. & Hurlle, J. M. Programmed cell death in the embryonic vertebrate limb. *Semin. Cell Dev. Biol.* **16**, 261–269 (2005).
- Fernández-Terán, M. A., Hinchliffe, J. R. & Ros, M. A. Birth and death of cells in limb development: a mapping study. *Dev. Dyn.* **235**, 2521–2537 (2006).
- Marazzi, G., Wang, Y. & Sassoon, D. Msx2 is a transcriptional regulator in the BMP4-mediated programmed cell death pathway. *Dev. Biol.* **186**, 127–138 (1997).
- Ferrari, D. et al. Ectopic expression of Msx-2 in posterior limb bud mesoderm impairs limb morphogenesis while inducing BMP-4 expression, inhibiting cell proliferation, and promoting apoptosis. *Dev. Biol.* **197**, 12–24 (1998).
- Pizette, S., Abate-Shen, C. & Niswander, L. BMP controls proximodistal outgrowth, via induction of the apical ectodermal ridge, and dorsoventral patterning in the vertebrate limb. *Development* **128**, 4463–4474 (2001).
- Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
- Mariani, F. V., Ahn, C. P. & Martin, G. R. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature* **453**, 401–405 (2008).
- Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
- Sanz-Ezquerro, J. J. & Tickle, C. Fgf signaling controls the number of phalanges and tip formation in developing digits. *Curr. Biol.* **13**, 1830–1836 (2003).
- Romer, A. S. *Vertebrate Paleontology* (Univ. Chicago Press, 1936).
- Prothero, D. R. & Foss, S. E. *The Evolution of Artiodactyls* (John Hopkins Univ. Press, 2007).
- Lopez-Rios, J. et al. Attenuated sensing of SHH by *Ptch1* underlies adaptive evolution of bovine limbs. *Nature* <http://dx.doi.org/10.1038/nature13289> (this issue).
- Chen, Y. & Struhl, G. Dual roles for Patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
- Li, Y., Zhang, H., Litington, Y. & Chiang, C. Cholesterol modification restricts the spread of Shh gradient in the limb bud. *Proc. Natl Acad. Sci. USA* **103**, 6548–6553 (2006).
- Butterfield, N. C. et al. Patched 1 is a crucial determinant of asymmetry and digit number in the vertebrate limb. *Development* **136**, 3515–3524 (2009).
- Sears, K. E. et al. Developmental basis of mammalian digit reduction: a case study in pigs. *Evol. Dev.* **13**, 533–541 (2011).
- Clifford, A. B. The evolution of the unguligrade manus in artiodactyls. *J. Vertebr. Paleontol.* **30**, 1827–1839 (2010).
- Cooper, L. N., Berta, A., Dawson, S. D. & Reidenberg, J. S. Evolution of hyperphalangy and digit reduction in the cetacean manus. *Anat. Rec. Adv. Integr. Anat. Evol. Biol.* **290**, 654–672 (2007).
- Rose, K. D. Skeleton of *Diacodexis*, oldest known artiodactyl. *Science* **216**, 621–623 (1982).
- Rose, K. D. On the origin of the order Artiodactyla. *Proc. Natl Acad. Sci. USA* **93**, 1705–1709 (1996).
- Theodor, J., Erfurt, J. & Metais, G. The earliest Artiodactyls: Diacodexidae, Dichobunidae, Homacodontidae, Leptochoeridae, and Raoellidae in *The Evolution of Artiodactyls* (eds. Prothero, D.R. & Foss, S.E.) (John Hopkins Univ. Press, 2007).
- Rasweiler, J. J., Cretokos, C. J. & Behringer, R. R. Alcian blue staining of cartilage of short-tailed fruit bat (*Carollia perspicillata*). *Cold Spring Harb. Protoc.* **2009**, pdb.p05165 (2009).
- Ovchinnikov, D. Alcian blue/alizarin red staining of cartilage and bone in mouse. *Cold Spring Harb. Protoc.* **2009**, pdb.p05170 (2009).
- Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
- Rasweiler, J. J., Cretokos, C. J. & Behringer, R. R. Whole-mount *in situ* hybridization of short-tailed fruit bat (*Carollia perspicillata*) embryos with RNA probes. *Cold Spring Harb. Protoc.* **2009**, pdb.p05164 (2009).
- Meredith, R. W. et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).

Acknowledgements We thank J. Lopez-Rios and R. Zeller (University of Basel, Switzerland) for generously providing data and discussion before publication. We also thank J. Carlos Izpisua Belmonte and A. Aguirre for sharing space and materials to complete experiments subsequent to review. Jerboa embryos were harvested with the assistance of S. Wu and colleagues in Xinjiang, China. Pig embryos were harvested with the assistance of D. Urban. Additional horse embryos were provided by R. Turner and H. Galatino-Homer (University of Pennsylvania) and by R. Fritsche and S. Lyle (Louisiana State University). Mouse *Gli1* probe plasmid, used in the pig *in situ*, was provided by A. Joyner. This work was supported by NIH grant R37HD032443 to C.J.T., and NSF IOS grant 1257873 to K.E.S.

Author Contributions K.L.C., K.E.S. and C.J.T. conceived of and initiated the project. K.L.C. and C.J.T. wrote the manuscript. K.L.C. performed the mouse, three- and five-toed jerboa, horse and camel *in situ* hybridizations, PH3 IHC, and skeletal stains. A.U. performed TUNEL and Sox9 IHC. J.M. and K.E.S. performed the pig *in situ* hybridizations. K.-S.B. cloned the pig probes. M.B. and D.A. provided most of the horse embryos and material for cloning the horse probes. J.A.S. provided the camel embryos and material for cloning the camel probes.

Author Information The probe sequence data for all genes and species has been deposited in the NCBI Probes database with the following accession numbers: CAMELBMP4, Pr032067180; CAMELFGF8, Pr032067181; CAMELGLI1, Pr032067182; CAMELHOXD13, Pr032067183; CAMELPTCH1, Pr032067184; HORSEBMP4, Pr032067185; HORSEFGF8, Pr032067186; HORSEGLI1, Pr032067187; HORSEHOXD13, Pr032067188; HORSEMSX2, Pr032067189; HORSEPTCH1, Pr032067190; HORSESHH, Pr032067191; JERBOABMP4, Pr032067192; JERBOAFGF8, Pr032067193; JERBOAGLI1, Pr032067194; JERBOAHOXD13, Pr032067195; JERBOAMSX2, Pr032067196; JERBOAPTCH1, Pr032067197; JERBOASHH, Pr032067198; MOUSEBMP4, Pr032067199; MOUSEFGF8, Pr032067200; MOUSEGLI1, Pr032067201; MOUSEHOXD13, Pr032067202; MOUSEMSX2, Pr032067203; MOUSEPTCH1, Pr032067204; MOUSESHH, Pr032067205; PIGFGF8, Pr032067206; PIGPTCH1, Pr032067207; PIGSHH, Pr032067208. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.L.C. (kcooper@ucsd.edu); correspondence and requests for pig embryos and materials should be addressed to K.E.S. (ksears@life.illinois.edu).

Attenuated sensing of SHH by *Ptch1* underlies evolution of bovine limbs

Javier Lopez-Rios^{1*}, Amandine Duchesne^{1,2*}, Dario Speziale¹, Guillaume Andrey³, Kevin A. Peterson⁴, Philipp Germann⁵, Erkan Ünal^{1,5}, Jing Liu⁴, Sandrine Floriot², Sarah Barbey⁶, Yves Gallard⁶, Magdalena Müller-Gerbl⁷, Andrew D. Courtney⁸, Christophe Klopp⁹, Sabrina Rodriguez^{2†}, Robert Ivanek^{1,10}, Christian Beisel⁵, Carol Wicking⁸, Dagmar Iber⁵, Benoit Robert¹¹, Andrew P. McMahon⁴, Denis Duboule^{3,12} & Rolf Zeller¹

The large spectrum of limb morphologies reflects the wide evolutionary diversification of the basic pentadactyl pattern in tetrapods. In even-toed ungulates (artiodactyls, including cattle), limbs are adapted for running as a consequence of progressive reduction of their distal skeleton to symmetrical and elongated middle digits with hoofed phalanges. Here we analyse bovine embryos to establish that polarized gene expression is progressively lost during limb development in comparison to the mouse. Notably, the transcriptional upregulation of the *Ptch1* gene, which encodes a Sonic hedgehog (SHH) receptor, is disrupted specifically in the bovine limb bud mesenchyme. This is due to evolutionary alteration of a *Ptch1* cis-regulatory module, which no longer responds to graded SHH signalling during bovine handplate development. Our study provides a molecular explanation for the loss of digit asymmetry in bovine limb buds and suggests that modifications affecting the *Ptch1* cis-regulatory landscape have contributed to evolutionary diversification of artiodactyl limbs.

A prime example of morphological evolution is the diversification of pentadactylous (five digits) limbs in tetrapods to serve a multitude of specialized needs such as grasping, digging, swimming, flying and running¹. Artiodactyls (even-toed ungulates, including cattle, deer, giraffe, hippopotamus, whale, dolphin, pig and camel) are a diversified order of mammals that underwent highly successful radiation from the early Eocene; that is, from ~55 million years ago^{2,3}. In the order Artiodactyla, Bovidae represent the largest extant family, and their domestication was pivotal to human agricultural settlement⁴. One key feature of terrestrial artiodactyls is the significant lengthening and streamlining of their distal limb skeleton, an adaption to walking and running on different terrains. Artiodactyl limbs are paraxonic, which means that the weight-bearing axis has shifted to between digits III and IV. The major morphological feature is that both middle digits are of equal length to facilitate walking on the distal-most phalanges (unguligrade posture). These evolutionary modifications in digit length and morphology are paralleled by varying degrees of digit loss (oligodactyly) and secondary fusions^{2,3,5}.

Analysis of *Bos taurus* (referred to as bovine) and *Mus musculus* (mouse) at equivalent developmental stages (Extended Data Fig. 1a) reveals the marked differences in their distal limb skeletons. In contrast to the mouse, only two elongated metacarpal primordia form in bovine handplates (autopod; Fig. 1a) and give rise to the cannon bone (Extended Data Fig. 1b)^{2,5}. The middle digits that give rise to the cloven hoof are symmetrical and of identical length (digits III and IV). The vestigial lateral condensations (digits II and V) form the dewclaws and the anterior-most digit I is absent (Fig. 1a), as in many other tetrapods^{6,7}. The oligodactyly of the bovine autopod could result from elimination of digit progenitors, but no apoptosis of cartilage precursors was detected

(Extended Data Fig. 2a). This excludes cell death as the probable cause of the oligodactyly, in agreement with previous analysis of pig limb buds, which represent Suidae, another major artiodactyl family⁸.

To circumvent pleiotropic and possibly lethal phenotypes, alterations underlying the unguligrade nature of the bovine limb skeleton should be limb-bud specific and probably alter the robust signalling systems that control vertebrate limb development^{9–12}. Anterior–posterior polarization of the nascent limb bud mesenchyme, activation of *Shh* expression, and upregulation of the SHH receptor *Ptch1* and the *Gli1* transcriptional regulator in responding mesenchymal cells occur without apparent differences in both mouse and bovine forelimb buds (Fig. 1b and Extended Data Fig. 2b)^{13–15}. In both species, expression of the extracellular BMP antagonist *Grem1* is upregulated in mesenchymal cells in response to SHH and keeps BMP activity low (Fig. 1b and Extended Data Fig. 2b)⁹. GREM1-mediated antagonism of BMPs contributes to maintaining *Fgf* expression in the apical ectodermal ridge (AER; Fig. 1b) as part of an epithelial–mesenchymal feedback loop¹¹. This analysis shows that the mesenchymal SHH and AER–FGF signalling centres and mesenchymal SHH targets are activated in an identical manner in limb buds of both species^{16,17}. Remarkably, expression domains scale perfectly in bovine limb buds, which are ~1.8-fold larger than mouse limb buds.

Molecular anterior–posterior limb-bud asymmetry loss

The first molecular differences become apparent as the posterior *Grem1* expression domain expands anteriorly in the mouse autopod (embryonic day (E) ~10.75). At equivalent bovine stages (D32), the shape of the *Grem1* domain appears symmetrical and distalized in bovine forelimb

¹Developmental Genetics, Department Biomedicine, University of Basel, CH-4058 Basel, Switzerland. ²Institut National de la Recherche Agronomique, Génétique Animale et Biologie Intégrative, F-78350 Jouy-en-Josas, France. ³School of Life Sciences, Federal Institute of Technology Lausanne, CH-1015 Lausanne, Switzerland. ⁴Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. ⁵Department for Biosystems Science and Engineering, Federal Institute of Technology Zurich and Swiss Institute of Bioinformatics, CH-4058 Basel, Switzerland. ⁶Institut National de la Recherche Agronomique, Domaine Expérimental du Pin au Haras, F-61310 Exmes, France. ⁷Institute of Anatomy, Department Biomedicine, University of Basel, CH-4056 Basel, Switzerland. ⁸Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia. ⁹Institut National de la Recherche Agronomique, Biométrie et Intelligence Artificielle, F-31326 Castanet-Tolosan, France. ¹⁰Swiss Institute of Bioinformatics, CH-4058 Basel, Switzerland. ¹¹Institut Pasteur, Génétique Moléculaire de la Morphogénèse and Centre National de la Recherche Scientifique URA-2578, F-75015 Paris, France. ¹²Department of Genetics and Evolution, University of Geneva, CH-1211 Geneva, Switzerland. †Present address: Institut National de la Recherche Agronomique, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, F-31077 Toulouse, France.

*These authors contributed equally to this work.

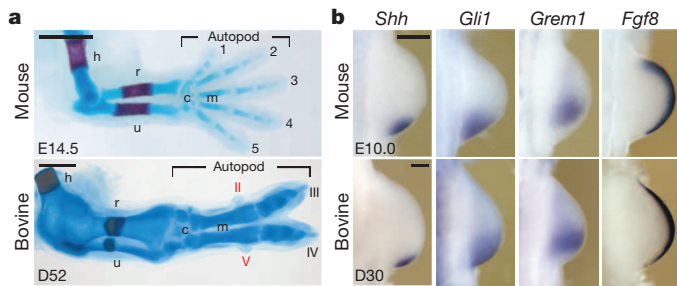


Figure 1 | Comparative analysis of mouse and bovine forelimb bud development. **a**, Forelimb skeletons at equivalent developmental stages in mouse (E14.5) and bovine (D52) embryos. 1–5 (mouse) and II–V (bovine), digits; c, carpus; h, humerus; m, metacarpus; r, radius; u, ulna. Vestigial digits II and V in bovine limbs are indicated in red. Scale bars, 1 mm. **b**, Comparative RNA *in situ* hybridization analysis during the onset of bovine and mouse forelimb bud development. Per gene and species, the transcript distribution was reproduced in at least $n = 3$ independent limb bud samples in independent experiments. In all figures, the images were scaled (scale bars, 0.25 mm) to enable direct comparison of both species and limb buds are always oriented with anterior to the top and posterior to the bottom.

buds, as illustrated by the reduction of the gap separating the dorsal and ventral *Grem1* domains (Fig. 2a). This distalization and loss of asymmetry is corroborated by *Hoxd13* expression in bovine forelimb buds (Fig. 2b and Extended Data Fig. 3a, b). These alterations are functionally relevant as the 5' *Hoxd* transcription factors, and in particular *Hoxd13*, regulate digit number and identity¹⁸. Concurrently, *Fgf8* expression in the AER is displaced distally and becomes symmetrical during bovine handplate development (Fig. 2c, d). AER FGFs regulate both limb-bud outgrowth and digit length during chondrogenesis^{19–21}. Therefore, the vestigial nature of digits II and V and loss of digit I in bovine can be explained by their territories getting precociously out of the AER FGF signalling range (Fig. 2d). All alterations in gene expression are limb-bud specific (Figs 2 and 3), as no differences were detected in other bovine tissues (Extended Data Fig. 3c).

Shh is essential for formation of digits 2–5 in the mouse and its expression in limb buds is controlled by a limb-specific *cis*-regulatory region^{16,22,23}. Genetic inactivation or evolutionary loss of this enhancer correlates with loss of limbs in different species^{24,25}. In lizards, the precocious down-regulation of *Shh* has been associated with digit loss²⁶. Therefore, we analysed the kinetics of *Shh* expression, but no marked differences between bovine and mouse were observed (Fig. 3a; see also Fig. 1b). As SHH acts in a graded manner to pattern the progenitors that give rise to the digits²⁷, we also assessed the SHH protein distribution (Fig. 3b). In mouse limb

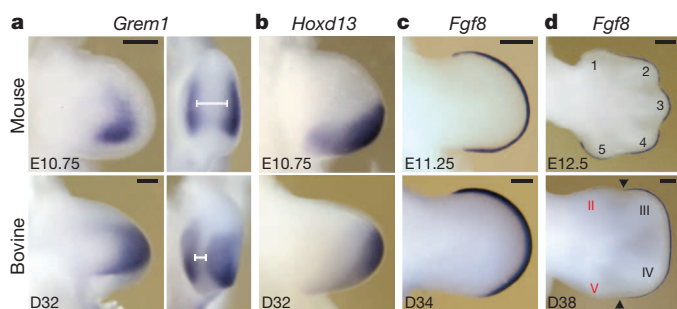


Figure 2 | Loss of molecular asymmetry in bovine forelimb buds. **a**, In contrast to the mouse, *Grem1* expression becomes more symmetrical (left, dorsal view) and distalized (right, apical view) in bovine forelimb buds by D32. The white line indicates the gap between the dorsal and ventral *Grem1* domains. **b**, Distal shift of *Hoxd13* expression in bovine forelimb buds. **c**, **d**, Analysis of *Fgf8* reveals the progressive distal shift of the AER in bovine limb buds, which becomes symmetrical at late stages (black arrowheads). Per gene and species, the transcript distribution was determined in at least $n = 3$ independent limb bud samples in independent experiments. Scale bars, 0.25 mm.

buds, the SHH protein is abundant in the posterior mesenchyme, whereas much lower levels are detected in the subapical mesenchyme (Fig. 3b, upper panels)²⁸. In contrast, SHH proteins extend beyond the apex in bovine limb buds (Fig. 3b, lower panels). Not only SHH protein levels but also SHH signalling were extended in bovine limb buds, as revealed by the anteriorly expanded expression of the direct transcriptional target *Gli1* (Fig. 3c).

Ptch1 alteration in the bovine limb bud

Interaction of SHH with its receptors PTCH1 and PTCH2 enables signal transduction, whereas unbound PTCH receptors inhibit the pathway. *Ptch* genes are direct targets of SHH and their transcriptional upregulation is highest close to the SHH source. This regulatory feedback enhances PTCH-mediated sequestering of ligands and effectively restricts the range of SHH diffusion^{29–31}. Therefore, the anterior expansion of the SHH protein and *Gli1* transcript domains could be a direct consequence of altered *Ptch* expression in bovine limb buds. In mouse limb buds, *Ptch1* is upregulated in the posterior mesenchyme and ectoderm³², whereas in bovine limb buds *Ptch1* is only upregulated in the ectoderm (Fig. 3d, e). The initially low mesenchymal *Ptch1* expression (Extended Data Fig. 4a) becomes undetectable during bovine autopod development (Fig. 3e). As *Ptch1* expression is very similar in all other bovine and mouse embryonic tissues (see for example Extended Data Fig. 3c), the alterations in upregulating *Ptch1* expression seem to be restricted to the bovine limb-bud mesenchyme. This mesenchymal *Ptch1* deficiency cannot be compensated by *Ptch2*, as its expression remains largely restricted to the ectoderm (Extended Data Fig. 4b). However, bovine limb-bud mesenchymal cells are able to respond to SHH as the expression of targets such as *Grem1*, *Hoxd13* and in particular *Gli1* (ref. 17) is upregulated (Figs 2a, b and 3c). To gain an insight into this regulatory conundrum, we derived a minimal network of experimentally verified interactions for mathematical simulations (Extended Data Fig. 5 and Supplementary Note). Removing the GLI-dependent *cis*-regulatory inputs on *Ptch1* expression from the simulations and fixing the rate of *Ptch1* production at low levels is sufficient to reproduce the anterior expansion of *Gli1* that is observed in bovine limb buds (Extended Data Fig. 5a–c, compare to Fig. 3c). Alterations of other parameters such as increasing SHH diffusion did not reproduce the opposing changes in *Gli1* and *Ptch1* expression (Extended Data Fig. 5d, compare to Fig. 3c, d).

A *Ptch1* limb bud *cis*-regulatory module

Molecular analysis and mathematical simulations indicated that alterations in the *Ptch1* *cis*-regulatory landscape could account for the failure to upregulate its mesenchymal expression in bovine limb buds. Previous analysis had pinpointed several *cis*-regulatory modules (CRMs) containing GLI binding sites around the transcriptional start site of the *Ptch1* locus^{17,33}. These CRMs overlap with regions of active chromatin as revealed by DNase I hypersensitivity and histone H3K27 acetylation (H3K27ac)³⁴ marks (Fig. 4a). Lentivector constructs³⁵ encoding the mouse or bovine *Ptch1* 5' proximal enhancer (PPE) direct expression of a *lacZ* reporter (with a β -globin minimal promoter; β *lacZ*) in a pattern similar to the endogenous gene with exception of limb buds (Fig. 4b)^{33,36}. Both PPE- β *lacZ* transgenes are expressed in the posterior AER (white arrowheads in Fig. 4b) but not in the mesenchyme, which is reminiscent of *Ptch1* expression in bovine limb buds (Fig. 3e). As the other mouse 5' CRMs identified by GLI1 chromatin immunoprecipitation (ChIP)³³ are not active in limb buds (K. A. Peterson and A. P. McMahon, unpublished), this indicated the possible existence of a mouse limb-bud mesenchymal CRM. Therefore, we used chromosome conformation capture in combination with deep sequencing (4C)^{37,38} to identify genomic regions that loop to the *Ptch1* proximal promoter in mouse limb buds (Fig. 4c). The 4C contact map reveals the preferential interaction of the *Ptch1* proximal promoter (used as bait) with a region located ~37 kilobases (kb) downstream. As this interaction was most prominent in posterior limb buds, we named this candidate region limb *cis*-regulatory module (LRM, Fig. 4c).

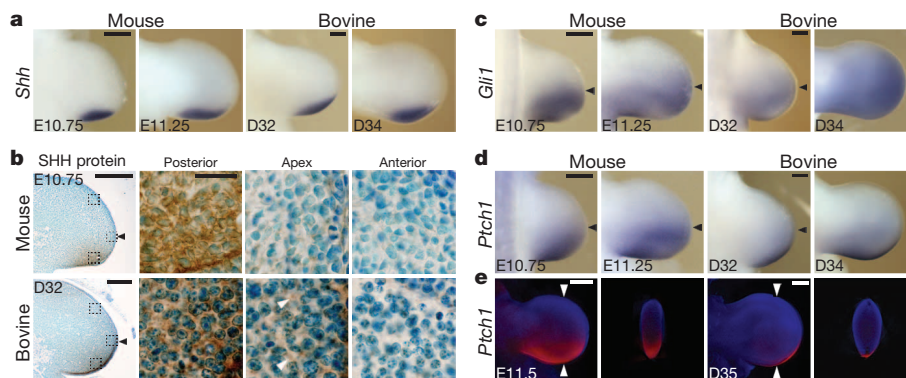


Figure 3 | Specific failure to upregulate mesenchymal *Ptch1* expression in bovine forelimb buds. **a**, *Shh* transcript distribution in mouse and bovine forelimb buds. **b**, SHH protein distribution (brown staining) on histological sections of forelimb buds. Nuclei are counterstained in blue. High-magnification insets are shown to the right (scale bar, 25 μ m). White arrowheads point to extracellular SHH immunocomplexes. **c**, *Gli1* transcription senses the range of SHH signalling. **d**, **e**, Analysis of *Ptch1*

expression by conventional (**d**) and optical projection tomography imaging (OPT, panel **e**). White arrowheads indicate the planes of OPT cross-sections shown to the right. Black arrowheads in **b–d** indicate the limb bud apex. Per gene and species, the transcript distribution was reproduced in at least $n = 3$ independent limb bud samples in independent experiments. Scale bars, 0.25 mm.

Inactivation of the bovine *Ptch1* LRM

The LRM is located in a genomic region between *Ptch1* coding exons 15–19 and includes several blocks of evolutionarily conserved intronic sequences that carry marks of active chromatin in mouse limb buds (Fig. 5a). As GLI proteins upregulate *Ptch1* expression in response to SHH signalling³³, their potential interactions with the LRM were studied by ChIP-seq analysis using a mouse strain with a 3 \times Flag epitope tag in the endogenous GLI3 protein. In mouse limb buds, GLI3 preferentially interacts with three of the evolutionarily conserved regions (Fig. 5a), and similar interactions were also detected for GLI2 and GLI1 proteins (data not shown). Together, these results indicated that at least two *cis*-regulatory regions could be present in the \sim 9-kb mouse LRM region (Fig. 5a). Comparison of LRM sequences from different mammalian species revealed two artiodactyl-specific insertions in these regions (red triangles, Fig. 5a). One of these is a microsatellite expansion, which varies considerably among different artiodactyls and is not present in non-artiodactyl mammals (Extended Data Fig. 6a). Therefore, the two conserved

regions of the mouse and bovine LRM (regions A and B, Fig. 5a) were cloned in tandem and assayed for enhancer activity in transgenic mouse embryos. The analysis of several independent founders showed that both *lacZ* transgenes are only expressed in the limb-bud core mesenchyme; that is, do not reproduce *Ptch1* expression (Extended Data Fig. 6b–d). This indicated that the non-conserved regions might be required for full LRM activity. Indeed, a conventional transgene encompassing the entire mouse LRM results in robust *lacZ* expression in the posterior limb bud and autopod mesenchyme, markedly similar to *lacZ* driven by the *Ptch1* genomic landscape (Fig. 5b, c and Extended Data Fig. 7a). In particular, the mouse LRM is activated by \sim E10.75 (Extended Data Fig. 7b); that is, around the developmental period when asymmetry is lost in bovine limb buds (Figs 2 and 3). Therefore, we assessed the transcriptional activity of the orthologous bovine LRM region. In contrast to the mouse, the bovine LRM-*lacZ* transgene remains restricted to the core mesenchyme (Fig. 5d and Extended Data Fig. 7c). As SHH signal transduction is intact in transgenic mouse limb buds, the bovine

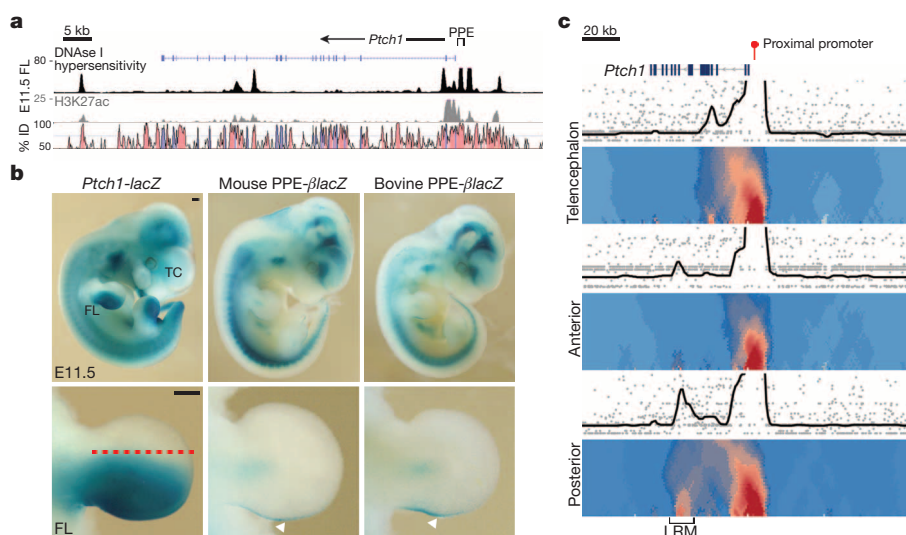


Figure 4 | Identification of a *Ptch1* limb *cis*-regulatory module (LRM). **a**, Scheme of the *Ptch1* locus with the profiles of DNase I hypersensitivity and H3K27ac marks³⁴ in mouse forelimb buds at E11.5. VISTA conservation plot comparison of the mouse and bovine *Ptch1* loci (mm9: chr13 63,592,959–63,682,958; see <http://genome.ucsc.edu>). **b**, Expression of a *lacZ* reporter inserted into the *Ptch1* locus compared with expression driven by the mouse and bovine 5' proximal enhancer (PPE, mouse: $n = 19$ of 32; bovine: $n = 24$ of

46 embryos). White arrowheads point to AER expression. FL, forelimb bud. Scale bars, 0.25 mm. Broken red line in panel **b** indicates the dissection of limb buds into anterior and posterior halves as used for 4C analysis. **c**, 4C contact analysis (mm9: chr13 63,575,000–63,760,000) using the proximal promoter as viewpoint, which reveals its interaction with an intragenic region in posterior limb buds. This interaction was not detected in the telencephalon (TC), which served as a control for the 4C analysis.

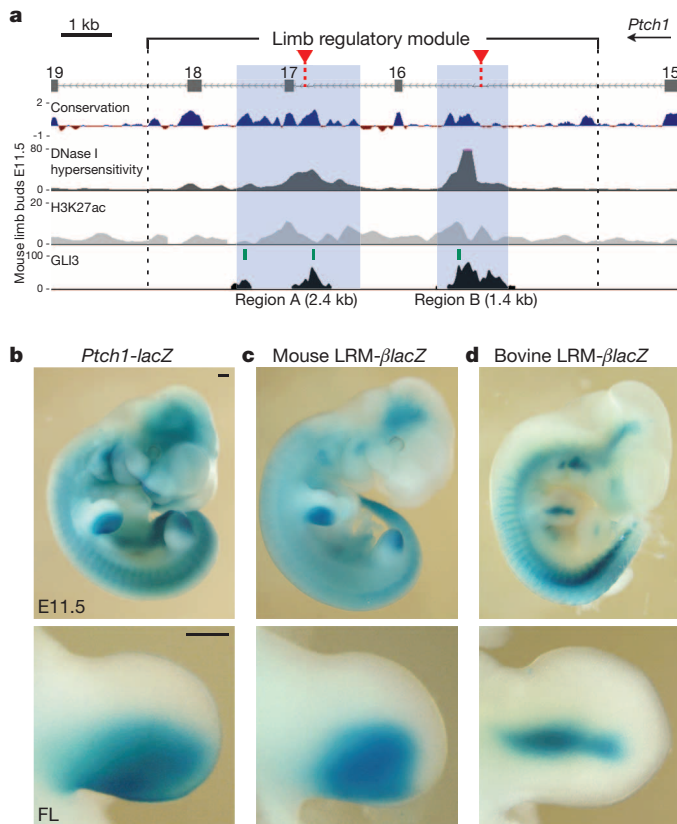


Figure 5 | Functional analysis of the mouse and bovine LRM in transgenic mouse embryos. **a**, Scheme depicting the mouse LRM in the *Ptch1* locus (mm9: chr13 63,621,577–63,634,154) with the placental mammal conservation plot (PhyloP; conserved regions in blue; fast-evolving regions in red), the distributions of DNase I hypersensitivity and H3K27ac marks, and the regions interacting with GLI3 proteins. Red triangles, artiodactyl-specific insertions. Green lines, conserved GLI binding motifs. Blue shaded areas indicate the conserved core regions A and B. **b–d**, Expression of a reporter *lacZ* transgene under control of the endogenous *Ptch1* locus (**b**), the mouse LRM (**c**) and the bovine LRM (**d**) in transgenic mouse embryos (mouse: $n = 4$ of 4; bovine: $n = 7$ of 12 embryos, others show no expression). FL, forelimb. Scale bars, 0.25 mm.

LRM is apparently unable to respond to graded SHH signalling (Fig. 5d, compare to Fig. 5b, c). Our transgenic analysis shows that (1) conserved and non-conserved regions are required for full activity of the mouse LRM, and (2) loss of its activity underlies the failure to upregulate *Ptch1* during bovine autopod development.

Prx1-Cre-mediated inactivation of *Ptch1* in the mouse limb-bud mesenchyme (*Ptch1*^{Δ/Δ}) causes loss of anterior–posterior asymmetry and predominant oligodactyly in forelimb buds³⁹, which resembles the alterations observed in bovine autopods. In *Ptch1*^{Δ/Δ} autopod primordia, *Gli1* expression expands even farther anterior than in bovine limb buds (Fig. 6a), whereas *Hoxd13* expression is distalized to a similar extent (Extended Data Fig. 8a). Most strikingly, the *Sox9* transcript distribution reveals the similarities in the pattern of digit condensations in bovine and mouse *Ptch1*^{Δ/Δ} autopods (Fig. 6b and Extended Data Fig. 8b; $n = 7$ of 18, others have only 3 condensations). In bovine and mouse *Ptch1*^{Δ/Δ} forelimb buds, the two lateral condensations are much reduced, whereas the primordia of the two central digits appear symmetrical and of equal length. This shows that loss of *Ptch1* is sufficient to phenocopy molecular and morphological hallmarks of the bovine autopod.

Discussion

Our study reveals the molecular changes underlying the loss of anterior–posterior asymmetry in bovine limb autopods. The symmetrical *Hoxd13*

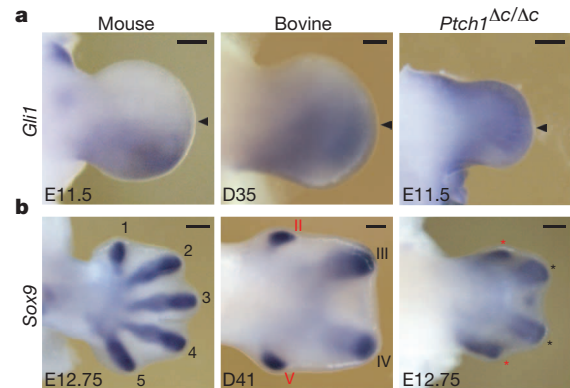


Figure 6 | Similarity of gene expression in bovine and mouse *Ptch1*^{Δ/Δ} forelimb buds. **a**, **b**, *Gli1* and *Sox9* expression in forelimb buds of mouse, bovine and *Ptch1*^{Δ/Δ} embryos. Vestigial digits are indicated in red. Asterisks indicate digits with uncertain identities. Arrowheads indicate limb bud apex. Per gene, species and mouse genotype, the transcript distribution was reproduced in at least $n = 3$ independent limb bud samples in independent experiments. Scale bars, 0.25 mm.

expression in the mesenchyme and more restricted *Fgf8* expression in the AER underlie the preferential and isomorphic elongation of the middle digits in bovine limbs^{18–20}. The distal restriction of AER *Fgf8* expression in bovine limb buds could be a consequence of differences in *Grem1* expression⁹ and/or increased SHH signalling to the AER³². The loss of autopod asymmetry is a likely consequence of the loss of activity of the LRM enhancer that normally upregulates *Ptch1* expression in the limb-bud mesenchyme in response to SHH signalling. During evolution, such a tissue-specific *cis*-regulatory rather than a pleiotropic alteration was probably favoured¹², as *Ptch1* inactivation causes embryonic lethality³⁶. Tissue-specific modifications in gene expression have previously been linked to the regulatory alterations underlying evolutionary diversification of embryonic structures^{24,40–43}. All these studies epitomize the potential for rewiring robust developmental pathways during morphological evolution^{44,45}.

Analysis of the mouse LRM indicates that this *cis*-regulatory module senses graded SHH signalling in the limb-bud mesenchyme and that both conserved and non-conserved regions are required for its full functionality. In agreement, active chromatin marks cover the entire mouse LRM and sequence analysis predicts additional species-specific GLI binding sites scattered throughout the non-conserved regions. Rather than being completely inactivated or deleted^{24,40}, the bovine LRM could have been altered in a way similar to the evolutionary diversification of the *ovo/shavenbaby* (*svb*) locus that controls formation of dorsal cuticular hairs in *Drosophila*^{42,46}. Alteration or loss of these structures in different *Drosophila* species requires multiple changes in the redundant and non-conserved *cis*-regulatory elements that control *svb* expression⁴⁶. These and other studies⁴¹ show how molecular alterations in several *cis*-regulatory elements with individually small phenotypic effects are able to change a robust trait and/or promote morphological convergence through parallel genetic changes under evolutionary constraints⁴⁷.

A similar mechanism might underlie the evolutionary diversification of artiodactyl autopods. The fossil records show that primitive artiodactyls were pentadactylous and reveal the progressive loss of anterior–posterior polarity in metacarpals and middle digits². This shift towards an unguligrade posture is paralleled by digit reductions and/or loss that occurred independently in different artiodactyl lineages⁴⁸. These findings support a progressive, and in part convergent, adaptation of the limb skeleton during artiodactyl evolution. Molecular analysis (this study and ref. 7) reveals the variability in limb bud development among different artiodactyl families, which is in agreement with progressive evolutionary adaptation. Similar to bovine, *Ptch1* expression is not upregulated in the mesenchyme of pig limb buds⁷. As Ruminantia (encompassing Bovidae) and Suina (encompassing Suidae) diverged early during artiodactyl

evolution^{2,3}, the mechanism that normally upregulates *Ptch1* in the autopod mesenchyme should have been disrupted by a common initiating alteration. However, *Ptch1* expression is upregulated in the limb-bud mesenchyme of camel embryos, which also diverged early during artiodactyl evolution. Another major difference is the elimination of lateral digit primordia by apoptosis in camel, but not bovine and pig limb buds (this study and refs 7,8). Our analysis provides evidence that the reduction and loss of digits in bovine limbs is secondary to the loss of autopod asymmetry caused by inactivation of the *Ptch1* LRM. One possible explanation for the differences in *Ptch1* expression⁷ could be that the alterations in the LRM that preceded divergence were compensated by genetic robustness involving other elements in certain artiodactyls such as camelids. Alternatively, the initiating events only weakened the inherent robustness of the *Ptch1* LRM in basal artiodactyls and subsequent inactivating alterations only occurred in some but not all lineages, which would explain why *Ptch1* upregulation is not disrupted in limb buds of all clades. Together with adaptive changes in other genes/pathways, these alterations would have caused the gradual loss of digit asymmetry and facilitated the adoption of unguligrade postures as evidenced by morphological analysis of extinct and modern artiodactyls.

METHODS SUMMARY

All experimental procedures involving bovine and mouse embryos were conducted in accordance with national laws, taking into account the 3R and the Basel Declaration principles.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 December 2013; accepted 27 March 2014.

Published online 18 June 2014.

- Polly, P. D. Limbs in mammalian evolution. In *Fins into Limbs* (Hall, B. K. ed.) 245–268 (Univ. Chicago Press, 2007).
- Prothero, D. R. & Foss, S. E. (eds) *The Evolution of Artiodactyls* (Johns Hopkins Univ. Press, 2007).
- Spaulding, M., O'Leary, M. A. & Gatesy, J. Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS ONE* **4**, e7062 (2009).
- Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl Acad. Sci. USA* **105**, 11597–11604 (2008).
- Budras, K. D. & Habel, R. E. (eds) *Bovine Anatomy* (Schlüttersche, 2011).
- de Bakker, M. A. et al. Digit loss in archosaur evolution and the interplay between selection and constraints. *Nature* **500**, 445–448 (2013).
- Cooper, K. L. et al. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* (in the press).
- Sears, K. E. et al. Developmental basis of mammalian digit reduction: a case study in pigs. *Evol. Dev.* **13**, 533–541 (2011).
- Benazet, J. D. et al. A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning. *Science* **323**, 1050–1053 (2009).
- Lopez-Rios, J. et al. GLI3 constrains digit number by controlling both progenitor proliferation and BMP-dependent exit to chondrogenesis. *Dev. Cell* **22**, 837–848 (2012).
- Zeller, R., Lopez-Rios, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nature Rev. Genet.* **10**, 845–858 (2009).
- Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
- Galli, A. et al. Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. *PLoS Genet.* **6**, e1000901 (2010).
- te Welscher, P., Fernandez-Teran, M., Ros, M. A. & Zeller, R. Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes Dev.* **16**, 421–426 (2002).
- Marigo, V., Johnson, R. L., Vortkamp, A. & Tabin, C. J. Sonic hedgehog differentially regulates expression of *GLI* and *GLI3* during limb development. *Dev. Biol.* **180**, 273–283 (1996).
- Zhu, J. et al. Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell* **14**, 624–632 (2008).
- Vokes, S. A., Ji, H., Wong, W. H. & McMahon, A. P. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev.* **22**, 2651–2663 (2008).
- Woltering, J. M. & Duboule, D. The origin of digits: expression patterns versus regulatory mechanisms. *Dev. Cell* **18**, 526–532 (2010).
- Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
- Mariani, F. V., Ahn, C. P. & Martin, G. R. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature* **453**, 401–405 (2008).
- Sanz-Ezquerro, J. J. & Tickle, C. Fgf signaling controls the number of phalanges and tip formation in developing digits. *Curr. Biol.* **13**, 1830–1836 (2003).
- Chiang, C. et al. Manifestation of the limb prepattern: limb development in the absence of Sonic Hedgehog Function. *Dev. Biol.* **236**, 421–435 (2001).
- Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
- Sagai, T. et al. Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). *Mamm. Genome* **15**, 23–34 (2004).
- Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
- Shapiro, M. D., Hanken, J. & Rosenthal, N. Developmental basis of evolutionary digit loss in the Australian lizard *Hemiergis*. *J. Exp. Zool. B* **297**, 48–56 (2003).
- Harfe, B. D. et al. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
- Gritli-Linde, A., Lewis, P., McMahon, A. P. & Linde, A. The whereabouts of a morphogen: direct evidence for short- and graded long-range activity of hedgehog signaling peptides. *Dev. Biol.* **236**, 364–386 (2001).
- Briscoe, J., Chen, Y., Jessell, T. M. & Struhl, G. A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* **7**, 1279–1291 (2001).
- Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing hedgehog. *Cell* **87**, 553–563 (1996).
- Marigo, V., Davey, R. A., Zuo, Y., Cunningham, J. M. & Tabin, C. J. Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996).
- Bouldin, C. M., Gritli-Linde, A., Ahn, S. & Harfe, B. D. Shh pathway activation is present and required within the vertebrate limb bud apical ectodermal ridge for normal autopod patterning. *Proc. Natl Acad. Sci. USA* **107**, 5489–5494 (2010).
- Vokes, S. A. et al. Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. *Development* **134**, 1977–1989 (2007).
- Cotney, J. et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* **22**, 1069–1080 (2012).
- Friedli, M. et al. A systematic enhancer screen using lentivector transgenesis identifies conserved and non-conserved functional elements at the Olig1 and Olig2 locus. *PLoS ONE* **5**, e15741 (2010).
- Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
- van de Werken, H. J. et al. 4C Technology: Protocols and Data Analysis. *Methods Enzymol.* **513**, 89–112 (2012).
- van de Werken, H. J. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9**, 969–972 (2012).
- Butterfield, N. C. et al. Patched 1 is a crucial determinant of asymmetry and digit number in the vertebrate limb. *Development* **136**, 3515–3524 (2009).
- Chan, Y. F. et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302–305 (2010).
- Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481–487 (2005).
- McGregor, A. P. et al. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
- Cretekos, C. J. et al. Regulatory divergence modifies limb length between mammals. *Genes Dev.* **22**, 141–151 (2008).
- Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).
- Peter, I. S. & Davidson, E. H. Evolution of gene regulatory networks controlling body plan development. *Cell* **144**, 970–985 (2011).
- Stern, D. L. & Frankel, N. The structure and evolution of cis-regulatory regions: the shavenbaby story. *Phil. Trans. R. Soc. B* **368**, 1632 (2013).
- Stern, D. L. The genetic causes of convergent evolution. *Nature Rev. Genet.* **14**, 751–764 (2013).
- Clifford, A. B. The evolution of the unguligrade manus in artiodactyls. *J. Vertebr. Paleontol.* **30**, 1827–1839 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Cooper and C. Tabin for sharing their results before publication and providing camel genomic DNA. Dolphin samples were received from C. Frere and D. Duffield. Pigmy hippopotamus mouth swabs for DNA extraction were provided by S. Furrer and B. Zimmerman from the Zoo of Zurich, whereas all other artiodactyl blood samples were donated from available frozen stocks by S. and M. Hoby from the Zoo of Basel. B. Wainwright provided the *Ptch1* conditional mouse line. We are grateful to E. Terszowska and A. Offinger for mouse care; V. Metis and M. Rondon for assistance in colony maintenance and collection of *Px1-Cre Ptch1^{Ac/Lac}* mouse embryos; and N. Dumesnil for assistance in collecting bovine embryos. All lentivector-mediated transgenic mouse embryos were produced by the EPFL transgenic platform, whereas conventional mouse transgenic embryos were produced commercially by Cyagen Biosciences Inc. We thank S. Beck-Cormier, A. Gritli-Linde, D. Haag-Wackernagel, Y. Lallemand, P. Zimmermann, G. Nusspaumer and A. Zuniga for technical advice and input and M. Bertaud, I. Ginez, D. Jarde and M. Moroldo for technical assistance. We are grateful to V. Taylor and members of our research groups for critical discussions and input on the manuscript. This research was supported by SNF grants 31003A_130803/146248 and the University of Basel (to R.Z.), EU

reintegration grant PERG-GA-2009-246576 (to J.L.-R.), SystemsX.ch iPhD Grant 20101078 (to D.I. and R.Z.), INRA (to A.D.), ANR grant No. 06-MRAR-027-01 (to B.R.), the EPFL and ERC grant SystemsHox.ch (to D.D.), NIH grant no. NS 033642 (to A.P.M.) and Australian National Health and Medical Research Council grant no. 569713 (to C.W.).

Author Contributions J.L.-R., A.D. and R.Z. conceived the project and wrote the manuscript; J.L.-R. and A.D. performed most of the mouse and bovine experimental studies. D.S. and E.U. performed mouse and bovine experiments under supervision of J.L.-R. and R.Z.; J.L.-R. and D.S. isolated, sequenced and analysed the ~9-kb LRM from different species. G.A. performed the lentivector studies under supervision of D.D. J.L. generated the *Gli3*^{3XFLAG} mouse strain and K.A.P. performed the GLI3 ChIP-seq analysis under supervision of A.P.M. P.G. performed the *in silico* simulations under supervision of D.I. S.F. participated in the design of bovine ISH probes and together with A.D., S.B. and Y.G. produced and collected the bovine embryos for the study. M.M.-G. did the skeletal analysis of bovine foetuses using CT scanning. A.D.C. and C.W. provided all

mouse embryos lacking *Ptch1* in the limb bud mesenchyme. C.B. performed all deep sequencing of the *Ptch1* loci from different species and as part of the 4C analysis. R.I. assisted in high-throughput data analysis. C.K. and S.R. participated in *in silico* assembly and analysis of the *Ptch1* loci from different species. B.R. was key to conceiving and initiating this project.

Author Information All 4C-seq data sets are deposited in the Gene Expression Omnibus repository under accession number GSE52988. The GLI3 ChIP-seq data relevant to this study are deposited in the Gene Expression Omnibus repository under accession number GSE52939. All the raw data for sequencing of the *Ptch1* locus are deposited in the European Nucleotide Archive (study accession number PRJEB5056). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.Z. (rolf.zeller@unibas.ch) or A.D. (amandine.duchesne@jouy.inra.fr, for bovine embryos and materials).

Structural basis for outer membrane lipopolysaccharide insertion

Haohao Dong^{1,2}, Quanju Xiang^{2,3}, Yinghong Gu¹, Zhongshan Wang^{1,2,4}, Neil G. Paterson⁵, Phillip J. Stansfeld⁶, Chuan He^{2,7}, Yizheng Zhang⁴, Wenjian Wang⁸ & Changjiang Dong¹

Lipopolysaccharide (LPS) is essential for most Gram-negative bacteria and has crucial roles in protection of the bacteria from harsh environments and toxic compounds, including antibiotics. Seven LPS transport proteins (that is, LptA–LptG) form a trans-envelope protein complex responsible for the transport of LPS from the inner membrane to the outer membrane, the mechanism for which is poorly understood. Here we report the first crystal structure of the unique integral membrane LPS translocon LptD–LptE complex. LptD forms a novel 26-stranded β -barrel, which is to our knowledge the largest β -barrel reported so far. LptE adopts a roll-like structure located inside the barrel of LptD to form an unprecedented two-protein ‘barrel and plug’ architecture. The structure, molecular dynamics simulations and functional assays suggest that the hydrophilic O-antigen and the core oligosaccharide of the LPS may pass through the barrel and the lipid A of the LPS may be inserted into the outer leaflet of the outer membrane through a lateral opening between strands β 1 and β 26 of LptD. These findings not only help us to understand important aspects of bacterial outer membrane biogenesis, but also have significant potential for the development of novel drugs against multi-drug resistant pathogenic bacteria.

All Gram-negative bacteria have an asymmetric outer membrane, in which the inner leaflet consists of phospholipid and the outer leaflet is comprised of lipopolysaccharide (LPS)^{1,2}. LPS normally has three components, lipid A, core oligosaccharide and O-antigen (Fig. 1a), which is matured at the periplasmic side of the inner membrane and is delivered to the outer membrane by seven LPS transport proteins^{3–7}. The LPS translocon LptD is unable to fold properly in the absence of LptE and the two proteins form a unique ‘barrel and plug’ architecture for LPS transport and insertion^{5,8,9}, the mechanism of which is poorly understood. LPS transport proteins are essential for most pathogenic Gram-negative bacteria and are highly conserved (Extended Data Fig. 1, 2). The LptD–LptE complex is a particularly attractive drug target¹⁰, as the drug candidates may not need to enter into the bacteria and would not be subject to efflux. The development of such drug candidates is hampered by the lack of a detailed model of the LptD–LptE complex. We report here the first X-ray crystallographic structure of the LptD–LptE complex, which provides the structural basis for LPS translocation across the outer membrane and insertion into the outer leaflet.

Structure of the LptD–LptE complex

The LptD model contained residues 226 to 786 with the LptE model comprising residues 19 to 169 (both numbering includes the signal peptide). LptD forms a 26 anti-parallel-stranded β -barrel with dimensions of approximately 70 Å in length, 50 Å in width and 50 Å in height (Fig. 1b–d, Extended Data Fig. 3a, b), whereas LptE forms a roll-like structure (Fig. 2a). Around three-quarters of LptE are located inside the β -barrel of LptD, the remainder extending into the periplasm (Extended Data Fig. 3c, d). The barrel is closed by strands β 1 and β 26 with the amino terminus in the periplasm and the carboxy terminus hidden inside the barrel on the periplasmic side (Fig. 1b, d). LptE is a lipoprotein with the acyl-lipid-modified amino terminus in the perfect position for insertion into the

outer membrane (Extended Data Fig. 3c, d). The largest previously reported single-protein barrels, PapC and FimD, comprise 24 anti-parallel β -strands and the barrels are completely blocked by a plug formed by the middle domain of PapC or FimD^{11,12}. In contrast, LptD forms a 26-stranded β -barrel and is ‘plugged’ by another protein, LptE. To our knowledge, LptD is the largest bacterial single-protein outer membrane β -barrel to date, and also the only barrel structure that has been demonstrated to use another protein as a plug^{5,13}.

To assess any structural changes in LptE owing to its formation of the LptD–LptE complex, we superimposed the LptE structure with the four available LptE crystal and NMR structures in the protein data bank (2R76, 3BF2, 4KWY and 2JXP). Although the amino acid identities are as low as 13%, the LptE structures are remarkably superimposable in terms of the α -helices and β -strands (Fig. 2b and Extended Data Fig. 4a–c). Given the essential roles of LptE in LptD folding and assembly, we predict that most of the LptD–LptE complexes of Gram-negative bacteria possess a similar structure to that of the complex that we present here.

Interactions of LptD and LptE

The β -strands in the LptD barrel are linked by loops, with those on the extracellular side much longer than those on the periplasmic face (Fig. 1b). Thirteen loops exist between strands at the extracellular side, namely Lp1 to Lp13. Most of these loops are at the surface of the pore. Lp4 and Lp8, however, are located in the interior of the barrel and consist of residues V334–Y354 and V519–S556, respectively (Fig. 1c and Extended Data Fig. 5a–e). The LptD–LptE structure confirms that these two loops are involved in LptD and LptE interactions^{5,9,14}. The surface area of monomeric LptE is 9,860.8 Å², of which 3,195.4 Å² is involved in formation of the LptD–LptE interface. Around 37 residues of LptE interact with 52 residues of LptD to stabilize the β -barrel (Extended Data Fig. 5a). Of particular note are the extensive interactions in the A87–T95 region of

¹Biomedical Research Centre, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. ²Biomedical Sciences Research Complex, School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST, UK. ³Department of Microbiology, College of Resource and Environment Science, Sichuan Agriculture University, Yaan 625000, China. ⁴College of Life Sciences, Sichuan University, Chengdu 610065, China. ⁵Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. ⁶Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ⁷School of Electronics and Information, Wuhan Technical College of Communications, No.6 Huangjiahu West Road, Hongshan District, Wuhan, Hubei 430065, China. ⁸Laboratory of Department of Surgery, the First Affiliated Hospital, Sun Yat-sen University, 58 Zhongshan Road II, Guangzhou, Guangdong 510080, China.

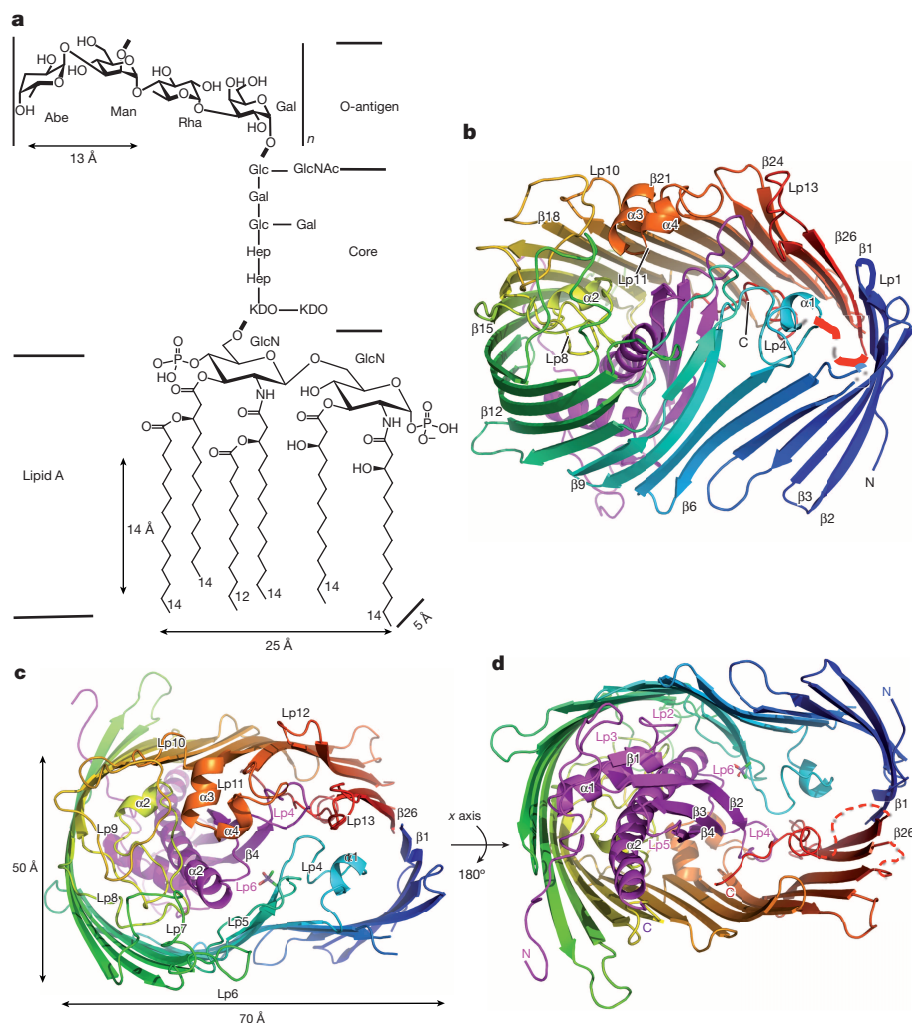


Figure 1 | Crystal structure of the LptD-LptE complex. The LptD barrel is shown in rainbow, the N-terminal strands coloured in blue, the C-terminal strands in red, and LptE in purple. **a**, Schematic representation of LPS. LPS contains three components and n ranges from 4 to 40. **b**, Cartoon representation of the LptD-LptE complex. The C-terminal domain of LptD forms the largest barrel identified to date, consisting of 26 anti-parallel β -strands. LptE is located inside the barrel. **c**, Top view of the LptD-LptE complex. Lp4 and Lp8 are located in the interior of the barrel and other loops at the surface of the barrel pore. **d**, Bottom view of the LptD-LptE complex.

LptE with Y680, Lp4 and the carboxyl-terminal residues T771–M786 of LptD (Fig. 2c and Extended Data Fig. 5b–e). In addition, LptD–LptE interactions are enhanced by hydrophobic interactions between amino-terminal residues W21–L23 of LptE and the outer surface of the LptD barrel (Fig. 2d).

To test the structural findings further, we generated single alanine or glycine amino acid substitution mutations, LptE deletions LptE(Δ W21–L23) and LptE(Δ A87–T95), and carboxyl-terminal truncations LptE(Δ T170–N196) and LptD(Δ T771–M786) (Extended Data Table 1a, b). Functional assays show that all the LptD and LptE mutants can grow in LB medium as well as their wild-types. Deletions LptE(Δ W21–L23), LptE(Δ A87–T95) and LptD(Δ T771–M786), however, significantly impair cell growth of *Escherichia coli* in LB medium containing 0.5% SDS and 1 mM EDTA, with the exception of LptE(Δ T170–N196), where the deleted residues are not involved in the LptD–LptE interaction (Fig. 2e, Extended Data Table 2a, b). Beside mutant LptE(Δ T170–N196), the other three deletion mutant proteins are expressed at similar levels to the wild-type proteins (Fig. 2f, g). We conclude that mutations of LptE(Δ W21–L23), LptE(Δ A87–T95) and LptD(Δ T771–M786) interfere with LptD–LptE interactions, either causing poor plugging of LptE in the LptD barrel or poor LptD assembly, thus increasing outer membrane permeability to SDS. These results are consistent with the structural findings and previous reports^{5,9,15–17}.

LptD structure suggests lateral opening

LptD forms a kidney-shaped pore with a 70 Å by 50 Å (outer side) and 50 Å by 30 Å (inner side) diameter at its widest points (Figs 1c, d and 3a, b). The extracellular loops of LptD seal most of the pore but leave a hole at one side of approximately 15 Å by 10 Å in diameter, which is

blocked by LptE to close the LptD channel completely (Fig. 3a, b). The lumen of the barrel is highly hydrophilic, similar to the polysaccharide translocators Wza¹⁸ and Alge¹⁹. How is the hydrophobic lipid A transported in the barrel of the LptD–LptE complex? LptE binds lipid A of LPS specifically¹⁴, indicating that LptE probably assists lipid A transport inside the LptD barrel. The hydrophobic residues inside the β -jellyroll structures of LptA and LptC have been identified to bind lipid A of LPS²⁰. We speculate that hydrophobic residues at the interface between the long $\alpha 2$ and the four β -strands of LptE may have a role in lipid A interaction inside the LptD barrel (Extended Data Figs 2, 4d). A study in *Neisseria meningitidis*, however, suggested that the LptE is not directly involved in LPS transport²¹. Therefore further investigation is required to clarify the role of LptE in LPS transport.

When viewed from the extracellular face, the β -strands of the barrel are twisted in an anticlockwise fashion relative to the periplasmic side (Fig. 1c, d). The angle of strands $\beta 1$ – $\beta 4$ is around 30° relative to the plane of the membrane, whereas that of strands $\beta 20$ – $\beta 26$ is about 67°, showing that strands $\beta 1$ and $\beta 26$ are partially separated at the periplasmic side (Fig. 3c, e, Extended Data Fig. 6a). This phenomenon has been seen in other outer membrane barrels FadL²², OmpW²³ and PagP²⁴, in which this feature laterally opens the barrel wall for hydrophobic substrate diffusion. There are only five hydrogen bonds between strands $\beta 1$ and $\beta 26$ of LptD (Fig. 3d), suggesting that lateral opening of the barrel between these strands is possible²⁵. Disulphide bond formation with LptD is essential for LPS translocation^{26,27} and this may have a role in any lateral opening of LptD. Intriguingly, C726 and C727 of LptD are located in the flexible loop between strands $\beta 24$ and $\beta 25$ at the periplasmic side (Fig. 3c). This loop is in a perfect position for disulphide

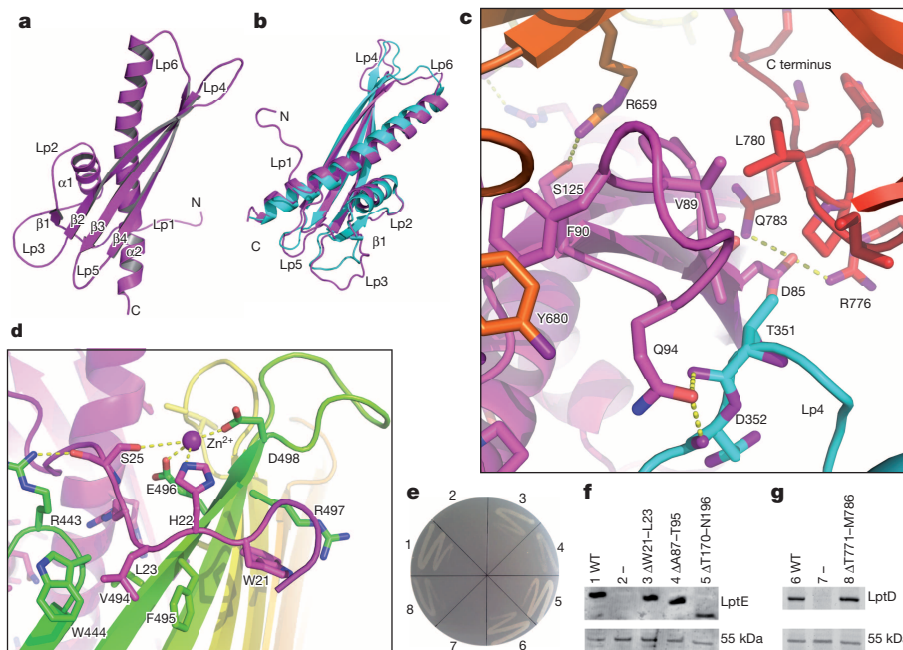


Figure 2 | The LptE and LptD interaction. LptD and LptE are coloured the same as in Fig. 1. **a**, The cartoon representation of the LptE structure from *S. typhimurium*. **b**, The LptE superimposes well with the LptE from *Shewanella oneidensis* (r.m.s.d. of 1.68 over 131 C α). LptE of *S. oneidensis* is shown in cyan. **c**, LptE interaction with residues on Lp4, Y680 and the C-terminal residues T771–M786 of LptD. **d**, LptE residues W21–L23 interact with hydrophobic residues at the outer side of the LptD barrel and a zinc ion mediates the LptD–LptE interaction. **e**, Truncation and deletion mutations of interaction residues of LptE and LptD impair cell growths in LB medium with 0.5% SDS and 1 mM EDTA with exception of LptE(Δ T170–N196). Lanes 1, 2, 3, 4 and 5 are from

bond formation with C31 and C173 in the N-terminal domain of LptD, which is directly linked to strand β 1 of LptD.

To test the hypothesis that a lateral opening of LptD is required for the LPS translocation, we generated single/double amino acid substitutions and deletion mutants of *lptD* (Extended Data Table 1b), and then performed functional assays. The deletion mutation LptD(Δ Q722–A729), removing C726 and C727, results in the death of the *E. coli* cells, consistent with previous reports^{26,27}. The structure of LptD suggests that the double amino acid mutation N232C/N757C may form a disulphide bond in the oxidative periplasm (Extended Data Fig. 6b), which would lock the strands β 1 and β 26 of LptD together and prevent any lateral opening. Indeed, the double mutation of N232C/N757C is lethal, whereas the single mutations of N232C and N757C and double mutations N232D/N757R and N232Y/N757H of LptD retain the same vitality as the wild-type (Fig. 3f, Extended Data Table 2b). The protein expression level of the double mutation N232C/N757C is similar to that of mutation N757C and N232Y/N757H in the cell membrane, strongly indicating that some degree of lateral opening between β 1 and β 26 of LptD is required for LPS translocation and insertion (Fig. 3g).

Simulations of the LptD–LptE gating

To further investigate the LPS translocation mechanism, molecular dynamics simulations of the LptD–LptE complex were performed. By simulating different lateral pressures, simultaneous separation of the strands β 1 and β 26 and opening of the LptD channel were observed when the pressures were below –65 bar (Extended Data Fig. 6c–e), with membrane lysis usually following within 10 ns. Although the applied pressures are non-physiological and are unlikely to resemble the true mechanism of gating in LptD–LptE, the strand separation does reveal that this is the weakest point within the structure and is therefore suggestive of the LPS insertion pathway. Pressures of –50 to –100 bar

AM689 cells with wild-type *lptE*, the empty pACYCDuet-1, *lptE*(Δ W21–L23), *lptE*(Δ A87–T95), *lptE*(Δ T170–N196), respectively, whereas 6, 7 and 8 are from AM661 cells with wild-type *lptD*, the empty pACYCDuet-1 and *lptD*(Δ T771–M786). **f**, The protein expression levels of the wild-type LptE and the deletion mutants on cell membrane are comparable with except of LptE(Δ T170–N196). An unidentified 55 kDa membrane protein was used as a loading control²⁷. –, Negative control (pACYCDuet-1 plasmid without *lptD* or *lptE* gene). **g**, The protein expression levels of wild-type LptD and LptD(Δ T771–M786) are similar. Panels **e**, **f** and **g** are representatives of three repeated experiments.

were also applied to simulations of other outer membrane proteins: BtuB (PDB code 2GUF) and FhaC (2QDZ) as well as the N232C/N757C mutant LptD–LptE complex. In these simulations no perturbation of the barrel architecture was observed before bilayer lysis.

Discussion and conclusion

Bacterial polysaccharides are transported across the outer membrane by either a Wza-like protein¹⁸, or via a β -barrel protein such as Alge¹⁹. The lumen of the LptD–LptE barrel is highly hydrophilic, a feature of translocases for hydrophilic polymer translocation^{18,19}. The O-antigen of lipopolysaccharide of *Salmonella typhimurium* consists of trisaccharide repeat units with short branches of single sugars (Fig. 1a), suggesting that the diameter of the linear polymer is around 13 Å for the O-antigen. Despite being occupied by LptE, the free cavity of LptD still has a diameter of 25 Å by 15 Å (Fig. 3b), which could easily accommodate the O-antigen. The O-antigen, which may consist of hundreds of saccharide units, has to pass through the LptD barrel, possibly using a similar mechanism to Alge¹⁹. The challenge of transporting large hydrophobic substrates through an aqueous barrel in the outer membrane is formidable. Nevertheless, several other outer membrane protein barrels adopt a lateral opening mechanism to diffuse the hydrophobic molecules to the outer membrane^{22–25}. LPS normally contains six fatty acyl chains, which would be 25 Å in length and 5 Å in width (Fig. 1a). Therefore, it would be big enough for the lipid chains of LPS to pass from the 5 Å side of lipid A to locate the LPS to the outer leaflet of the outer membrane if strands β 1 and β 26 could open around 7–9 Å. The diameter of the free cavity of the LptD barrel with LptE is 25 Å by 15 Å, which may help to position the 5 Å side of lipid A to the gate between the strands β 1 and β 26. Once the non-consecutive disulphide bonds form, the LptD N terminus links to LptA and forms a trans-envelope complex together with other LPS transport proteins from the inner membrane to

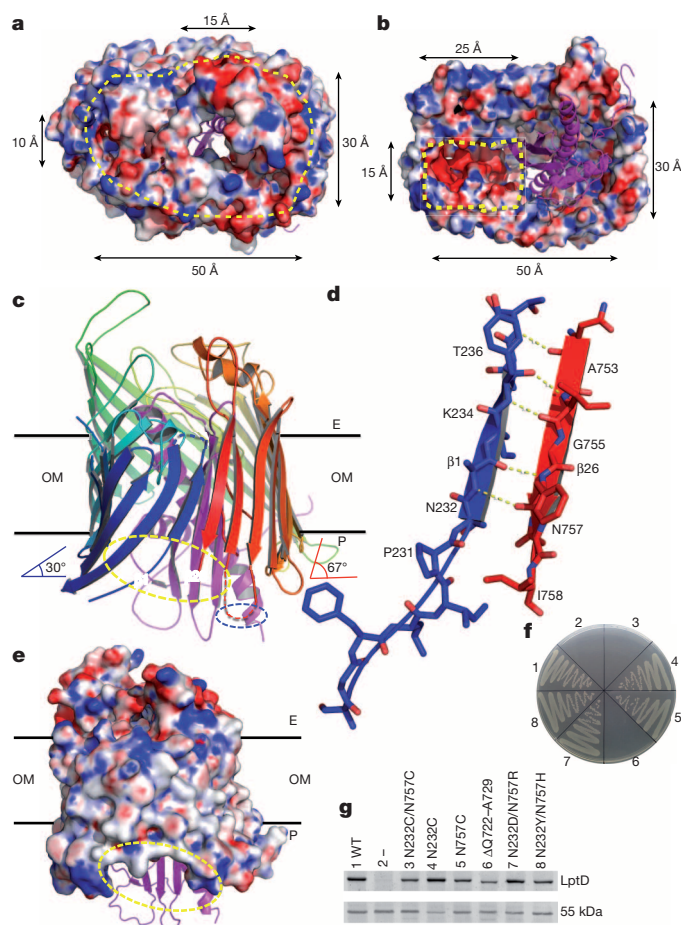


Figure 3 | The unique barrel and pore of the LptD-LptE complex. The colour of the cartoon representation is identical to Fig. 1. In the electrostatic surface potential map, negatively charged residues are shown in red and positively charged residues in blue. **a**, Top view of the electrostatic surface potential map of the LptD pore. The yellow dotted line shows the pore, which is covered by extracellular loops. LptE blocks a hole at one side of the pore. **b**, Bottom view of electrostatic surface potential map of the barrel. **c**, The twisted barrel. The tilt angle of the N-terminal strands is 30° and the tilt angle of the C-terminal strands is 67°. Strands β1 and β26 are separated at the periplasmic side, which is highlighted by a yellow dotted circle. The positions of C726 and C727 for disulphide bond formations are shown in a blue dotted circle. E, P and OM indicate extracellular side, periplasmic side and outer membrane, respectively. **d**, The interaction between strands β1 and β26. **e**, Side view of the LptD barrel, which shows the hydrophobic belt of the barrel exterior. The yellow dotted circle shows the separation between strands β1 and β26. **f**, Double cysteine mutation N232C/N757C, which potentially locks β1 and β26, causes the death of *E. coli*. Deletion LptD Δ Q722-A729 is lethal. 1, 2, 3, 4, 5, 6, 7 and 8 represent AM661 cells with the wild-type *lptD*, the empty pACYCDuet-1, *lptD* double mutation N232C/N757C, N232C, N757C, Δ Q722-A729, N232D/N757R and N232Y/N757H, respectively. **g**, The protein expression level of LptD mutant N232C/N757C is similar to that of LptD mutants N757C and N232Y/N757H on the cell membrane, but LptD mutant N232C/N757C could not survive possibly because it prevents the lateral opening. The unidentified 55-kDa membrane protein was used as a loading control²⁷. –, Negative control (pACYCDuet-1 plasmid without the *lptD* gene). Panels **f** and **g** are representatives of three repeated experiments.

the outer membrane. We propose that, by using ATP, the LPS is extracted from the inner membrane and passed to LptC²⁸, LptA²⁹ and the LptD-LptE complex sequentially³⁰, triggering a lateral opening of LptD and opening of the LptD channel and thus promoting the lateral insertion of LPS into the outer membrane through the gate between the strands β1 and β26 of LptD (Fig. 4). LPS is highly negatively charged with divalent cations bridging neighbouring LPS molecules to form the outer membrane

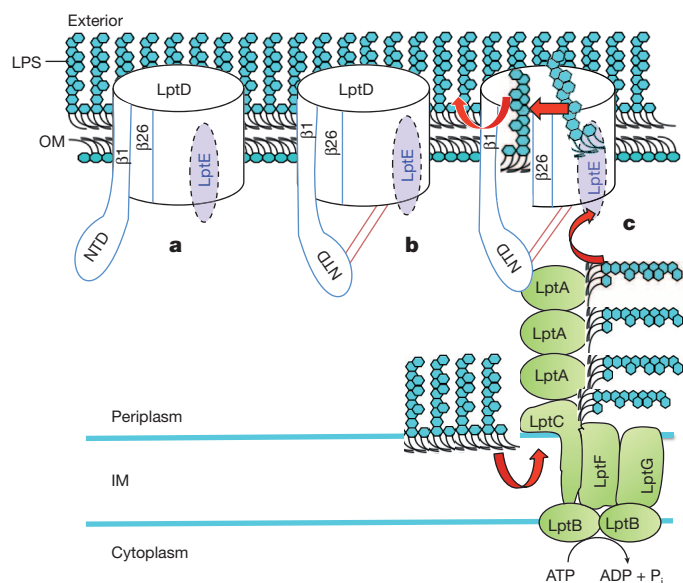


Figure 4 | The proposed mechanism of LPS transport. **a**, The newly synthesized LptD has not formed the disulphide bonds and the N-terminal domain of LptD is flexible. **b**, LptD forms the disulphide bonds (red lines), resulting in the conformational change of the N-terminal domain (NTD) of LptD. **c**, The N-terminal domain of LptD interacts with LptA, forming the seven-protein trans-envelope complex. LPS is highly negatively charged, the resulting electrostatic repulsion between LPS molecules probably separates them in the inner membrane (IM) in the periplasm. By using the ATP, the LPS molecules are extracted from the inner membrane by LptBCFG and passed on to LptC, LptA and the LptD-LptE complex, triggering the simultaneous opening of the LptD pore channel and the lateral gate for LPS transport and insertion. There is a high concentration of divalent cations at the outer leaflet of the outer membrane, and they are well known to bring neighbouring LPS molecules together. The electrostatic interaction would drive the newly arrived LPS molecules to insert lipid A precisely at the outer leaflet of the outer membrane.

permeability barrier. We expect that once LPS reaches the cation-rich outer leaflet of the outer membrane, forces of hydrophobic interaction between lipid A subunits of LPS and divalent-cation-mediated electrostatic interactions between LPS core oligosaccharide subunits would drive the insertion of LPS into the outer leaflet of the outer membrane. The mechanism of LptD-LptE lateral translocation of LPS differs from that of other lateral transport barrels^{22–25} in the way that LPS is delivered to the outer membrane external leaflet. The barrel of LptD is essentially a closed circle, despite the opening observed between strands β1 and β26, as the non-consecutive disulphide bonds link the N terminus of LptD and the periplasmic loop between strands β24 and β25 of LptD (Fig. 3c). The lateral opening between the strands β1 and β26 and the disulphide bonds would appear to ensure that LPS must enter the barrel and insert correctly into the outer leaflet of the outer membrane rather than the inner leaflet of the outer membrane.

METHODS SUMMARY

The *LptD* and *LptE* genes of *S. typhimurium* strain LT2 were cloned into pET-28b(+) and pACYCDuet-1, respectively, and the LptD-LptE complex was over-expressed using the two plasmids in *E. coli* C43(DE3). The LptD-LptE complex was purified using nickel-affinity and gel-filtration chromatography. The four-wavelength multi-wavelength anomalous dispersion (MAD) data sets were collected on beamline I24 at Diamond Light Source using selenomethionine-labelled crystals (Extended Data Fig. 7a–g, Extended Data Table 3). For functional assays of LptD and LptE, *E. coli* strain AM689, *E. coli* strain AM661 and pACYCDuet-1 plasmid were used. To test the LptD and LptE expression levels, hexa-His tag was introduced into the LptD after the signal peptide cleavage site and the C terminus of LptE, and the western blots were carried out using His-tag antibody. All molecular dynamics simulations of the LptD-LptE complex were performed using GROMACS v4.6.4.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 October 2013; accepted 12 May 2014.

Published online 18 June 2014.

1. Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).
2. Ruiz, N., Kahne, D. & Silhavy, T. J. Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature Rev. Microbiol.* **7**, 677–683 (2009).
3. Freinkman, E., Okuda, S., Ruiz, N. & Kahne, D. Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. *Biochemistry* **51**, 4800–4806 (2012).
4. Villa, R. *et al.* The *Escherichia coli* Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *J. Bacteriol.* **195**, 1100–1108 (2013).
5. Freinkman, E., Chng, S. S. & Kahne, D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proc. Natl Acad. Sci. USA* **108**, 2486–2491 (2011).
6. Sperandio, P. *et al.* Characterization of *lptA* and *lptB*, two essential genes implicated in lipopolysaccharide transport to the outer membrane of *Escherichia coli*. *J. Bacteriol.* **189**, 244–253 (2007).
7. Sperandio, P. *et al.* Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *J. Bacteriol.* **190**, 4460–4469 (2008).
8. Wu, T. *et al.* Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 11754–11759 (2006).
9. Chimalakonda, G. *et al.* Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108**, 2492–2497 (2011).
10. Srinivas, N. *et al.* Peptidomimetic antibiotics target outer-membrane biogenesis in *Pseudomonas aeruginosa*. *Science* **327**, 1010–1013 (2010).
11. Remaut, H. *et al.* Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* **133**, 640–652 (2008).
12. Phan, G. *et al.* Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate. *Nature* **474**, 49–53 (2011).
13. Fairman, J. W., Noinaj, N. & Buchanan, S. K. The structural biology of beta-barrel membrane proteins: a summary of recent reports. *Curr. Opin. Struct. Biol.* **21**, 523–531 (2011).
14. Chng, S. S., Ruiz, N., Chimalakonda, G., Silhavy, T. J. & Kahne, D. Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. *Proc. Natl Acad. Sci. USA* **107**, 5363–5368 (2010).
15. Grabowicz, M., Yeh, J. & Silhavy, T. J. Dominant negative *lptE* mutation that supports a role for LptE as a plug in the LptD barrel. *J. Bacteriol.* **195**, 1327–1334 (2013).
16. Braun, M. & Silhavy, T. J. Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol. Microbiol.* **45**, 1289–1302 (2002).
17. Ruiz, N., Falcone, B., Kahne, D. & Silhavy, T. J. Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell* **121**, 307–317 (2005).
18. Dong, C. *et al.* Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature* **444**, 226–229 (2006).
19. Whitney, J. C. *et al.* Structural basis for alginate secretion across the bacterial outer membrane. *Proc. Natl Acad. Sci. USA* **108**, 13083–13088 (2011).
20. Okuda, S., Freinkman, E. & Kahne, D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in *E. coli*. *Science* **338**, 1214–1217 (2012).
21. Bos, M. P. & Tommassen, J. The LptD chaperone LptE is not directly involved in lipopolysaccharide transport in *Neisseria meningitidis*. *J. Biol. Chem.* **286**, 28688–28696 (2011).
22. van den Berg, B., Black, P. N., Clemons, W. M., Jr & Rapoport, T. A. Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304**, 1506–1509 (2004).
23. Hong, H., Patel, D. R., Tamm, L. K. & van den Berg, B. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J. Biol. Chem.* **281**, 7568–7577 (2006).
24. Khan, M. A. & Bishop R. E. Molecular mechanism for the lateral lipid diffusion between the outer membrane external leaflet and a beta-barrel hydrocarbon ruler. *Biochemistry* **48**, 9745–9756 (2009).
25. Noinaj, N. *et al.* Structural insight into the biogenesis of β -barrel membrane proteins. *Nature* **501**, 385–390 (2013).
26. Chng, S. S. *et al.* Disulfide rearrangement triggered by translocon assembly controls lipopolysaccharide export. *Science* **337**, 1665–1668 (2012).
27. Ruiz, N., Chng, S. S., Hiniker, A., Kahne, D. & Silhavy, T. J. Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proc. Natl Acad. Sci. USA* **107**, 12245–12250 (2010).
28. Tran, A. X., Dong, C. & Whitfield, C. Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *J. Biol. Chem.* **285**, 33529–33539 (2010).
29. Suits, M. D., Sperandio, P., Deho, G., Polissi, A. & Jia, Z. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J. Mol. Biol.* **380**, 476–488 (2008).
30. Sperandio, P. *et al.* New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA–LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *J. Bacteriol.* **193**, 1042–1053 (2011).

Acknowledgements We thank T. J. Silhavy for providing AM689 and AM661 cells, J. Naismith and C. Whitfield for support, and R. Field, A. Johnston and X. Tang for critical reading of the manuscript. We thank the staff at I24, I02, I03 and I04 of Diamond Light Source UK for beam time (proposal mx7641) and their assistance with data collection. C.D. is a Wellcome Trust career development fellow (083501/Z/07/Z). Q.X. and Z.W. are the recipients of the Chinese overseas study scholarships of The China Scholarship Council.

Author Contributions C.D. and W.W. conceived and designed the experiments. H.D., Q.X., C.H. expressed, purified and crystallized the LptD–LptE complex, and Y.G. and Z.W. did the mutagenesis and the functional assays. C.D., N.G.P., H.D. and W.W. undertook data collection and structure determination. Q.X., Z.W. and Y.Z. generated the constructs for the protein expression and P.J.S. performed the molecular dynamics simulation. C.D., H.D., W.W. and N.G.P. wrote the manuscript.

Author Information The atomic coordinate and the structure factor of the LptD–LptE is deposited at the Protein Data Bank under access code 4N4R. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.D. (c.dong@uea.ac.uk) or W.W. (wenjian166@gmail.com).

A close-pair binary in a distant triple supermassive black hole system

R. P. Deane^{1,2}, Z. Paragi³, M. J. Jarvis^{4,5}, M. Coriat^{1,2}, G. Bernardi^{2,6,7}, R. P. Fender⁴, S. Frey⁸, I. Heywood^{6,9}, H.-R. Klöckner¹⁰, K. Grainge¹¹ & C. Rumsey¹²

Galaxies are believed to evolve through merging¹, which should lead to some hosting multiple supermassive black holes^{2–4}. There are four known triple black hole systems^{5–8}, with the closest black hole pair being 2.4 kiloparsecs apart (the third component in this system is at 3 kiloparsecs)⁷, which is far from the gravitational sphere of influence (about 100 parsecs for a black hole with mass one billion times that of the Sun). Previous searches for compact black hole systems concluded that they were rare⁹, with the tightest binary system having a separation of 7 parsecs (ref. 10). Here we report observations of a triple black hole system at redshift $z = 0.39$, with the closest pair separated by about 140 parsecs and significantly more distant from Earth than any other known binary of comparable orbital separation. The effect of the tight pair is to introduce a rotationally symmetric helical modulation on the structure of the large-scale radio jets, which provides a useful way to search for other tight pairs without needing extremely high resolution observations. As we found this tight pair after searching only six galaxies, we conclude that tight pairs are more common than hitherto believed, which is an important observational constraint for low-frequency gravitational wave experiments^{11,12}.

SDSS J150243.09+111557.3 (J1502+1115 hereafter) was identified as a quasar at $z = 0.39$ with double-peaked [O III] emission lines, which can be a signature of dual active galactic nuclei (AGN)¹³. Adaptive optics assisted near-infrared (K-band) images reveal two components separated by 1.4 arcsec (7.4 kpc) and are referred to as J1502P and J1502S¹⁴. Integral-field spectroscopy determined that the double-peaked [O III] emission in the SDSS spectrum is explained by a 657 km s^{-1} velocity offset between J1502S and J1502P, which are dust-obscured and unobscured AGN, respectively. These two components were also observed as steep-spectrum radio sources ($\alpha < -0.8$, with flux density $S_\nu \propto \nu^2$, where ν is the emission frequency) with the Jansky Very Large Array (JVLA) at 1.4, 5 and 8 GHz. The combination of the above results supported the discovery of a kiloparsec-scale dual AGN system¹⁵.

We performed 1.7 GHz and 5 GHz Very Long Baseline Interferometry (VLBI) observations of J1502+1115 with the European VLBI Network (EVN), revealing two flat-spectrum ($\alpha \approx -0.1 \pm 0.1$) radio components within the spatially unresolved near-infrared component J1502S. Flat radio spectra identify optically thick radio emission, which is characteristic of the core of relativistic jets generated by an accreting black hole¹⁶. The two components (labelled J1502SE and J1502SW in Fig. 1a) are each marginally resolved and have an angular separation of $\sim 26 \text{ mas}$, which corresponds to a projected spatial separation of $\sim 138 \text{ pc}$ at $z = 0.39$. Both J1502SE and J1502SW have radio luminosities of $L_{1.7} = 7 \times 10^{23} \text{ W Hz}^{-1}$ and brightness temperatures of $T_B \gtrsim 2 \times 10^8 \text{ K}$, consistent with actively accreting, intermediate radio luminosity nuclei. The high brightness temperatures, flat spectra, and co-spatial centroids (of both J1502SE and J1502SW) at both frequencies strongly support that

these two radio components are associated with two separate, accreting supermassive black holes (SMBHs). Alternative scenarios are discussed and ruled out in the Methods section. Arcminute Microkelvin Imager (AMI) Large Array observations at 15.7 GHz suggest that the two radio cores (J1502SE/SW) flatten the overall J1502S spectrum at higher frequency (Extended Data Fig. 1). The third active nucleus (J1502P) at a projected distance of 7.4 kpc from J1502S has a steep radio spectrum and is not detected in the VLBI observations.

A re-analysis of archival JVLA observations provides supporting evidence for the inner binary discovery in J1502S. The JVLA 5 GHz map of J1502S and J1502P is shown in Fig. 1b, and Fig. 1c shows the point-source-subtracted JVLA 5 GHz residual map. The latter reveals 'S'-shaped radio emission detected at high significance and which is rotationally symmetric about the flat-spectrum nuclei (a larger version of Fig. 1c is shown in Extended Data Fig. 2). The 5 GHz inner-jet axis is offset by $\sim 45^\circ$ from the position angle of the vector between J1502SE and J1502SW (Extended Data Fig. 3). The location and close proximity of the two flat-spectrum nuclei support the view that the 'S'-shaped radio emission is a modulation (or bending) of one pair of jets associated with one of the nuclear components. This type of 'S'-shaped structure is commonly attributed to precessing jets, with a resulting radio-jet morphology similar to the famous X-ray binary SS 433¹⁷. Jet precession in AGN has long been predicted to be associated with the presence of binary black holes³. The spatially resolved detection of the inner binary centred on the 'S'-shaped jet emission provides, for the first time, direct evidence of this prediction. It therefore demonstrates that this is a promising method to find close-pair binary SMBHs that cannot be spatially resolved by current instruments ($\ll 1 \text{ mas}$ angular resolution), yielding excellent targets for pointed gravitational wave experiments. The configuration where a primary black hole emits significant extended jet emission while the secondary appears to have none is also seen in the lowest-separation binary SMBH known, namely the VLBI-discovered system 0402+379 at $z \approx 0.06$ (ref. 10).

The host galaxies associated with J1502P and J1502S have elliptical galaxy morphologies and stellar masses of $M_* = (1.7 \pm 0.3) \times 10^{11} M_\odot$ and $(2.4 \pm 0.4) \times 10^{11} M_\odot$, respectively¹⁵, where M_\odot is the solar mass. J1502P has a measured black-hole mass (M_{BH}) given by $\log(M_{\text{BH}}/M_\odot) = 8.06 \pm 0.24$ (ref. 15), which is consistent with the black hole to bulge mass ($M_{\text{BH}}-M_{\text{bulge}}$) relation¹⁸. If the same is true for J1502S, this implies that J1502SE/SW both have M_{BH} values of $\sim 10^8 M_\odot$ and each have a sphere of influence of $\sim 10 \text{ pc}$ (the radius at which the black hole dominates the gravitational potential relative to the host galaxy). The J1502S optical spectrum reveals a single [O III] component, despite the double flat-spectrum cores. If a circular orbit of radius $a = 138/2 = 69 \text{ pc}$ and a black hole binary mass M_{12} of $2 \times 10^8 M_\odot$ are assumed, the maximum expected velocity offset between J1502SE/SW is $V_{\text{J1502E/W}} = \sqrt{GM_{12}/a}$

¹Astrophysics, Cosmology and Gravity Centre, Department of Astronomy, University of Cape Town, Rondebosch, 7701, Cape Town, South Africa. ²Square Kilometre Array South Africa, Pinelands, 7405, Cape Town, South Africa. ³Joint Institute for VLBI in Europe, 7990 AA, Dwingeloo, The Netherlands. ⁴Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK. ⁵Physics Department, University of the Western Cape, Bellville, 7535, South Africa. ⁶Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, Grahamstown, 6140, South Africa. ⁷Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. ⁸Satellite Geodetic Observatory, Institute of Geodesy, Cartography and Remote Sensing, H-1592, Budapest, Hungary. ⁹Australia Telescope National Facility, CSIRO Astronomy and Space Science, Epping, New South Wales 1710, Australia. ¹⁰Max-Planck-Institut für Radioastronomie, D-53121 Bonn, Germany. ¹¹Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester, UK. ¹²Astrophysics Group, Cavendish Laboratory, University of Cambridge, Cambridge, UK.

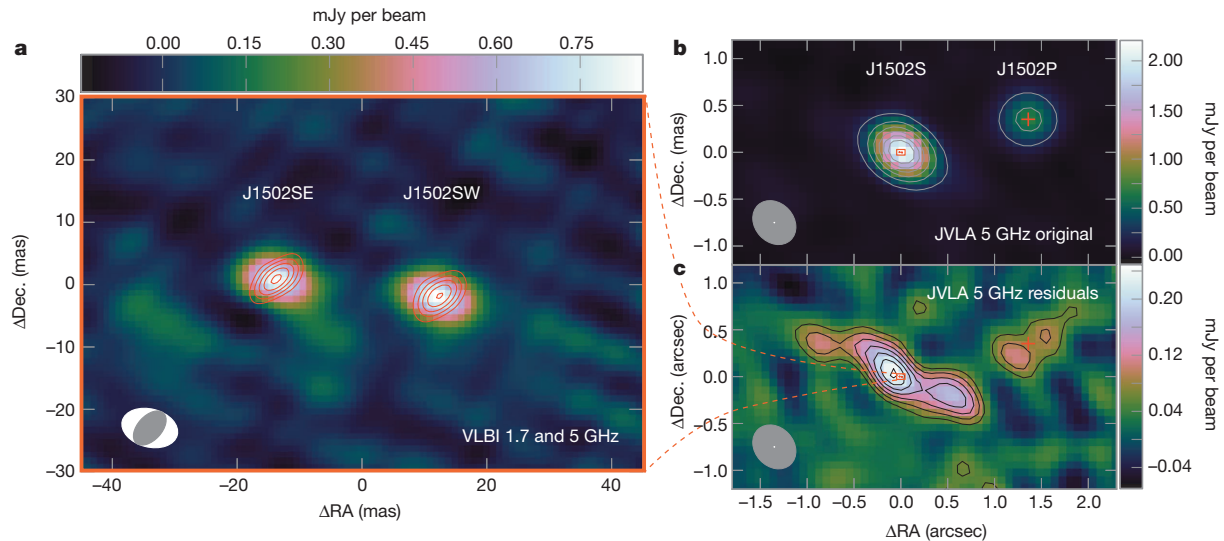


Figure 1 | VLBI and JVL A maps of the triple supermassive black hole system in J1502+1115. **a**, 5 GHz VLBI map (colour scale) showing the two components J1502SE and J1502SW. Over-plotted are red contours of the 1.7 GHz VLBI map with levels beginning at 400 μ Jy per beam and increasing in steps of 100 μ Jy per beam ($\sim 3\sigma$). RA, right ascension; dec., declination. **b**, JVL A 5 GHz map imaged with Briggs *uv*-weighting (robust = 1; see discussion on weighting in Methods). Over-plotted are the 5 GHz contours (grey) starting at 250 μ Jy per beam and increasing in steps of 300 μ Jy per beam (20σ). **c**, Point-source subtracted JVL A 5 GHz residuals, revealing a rotationally symmetric

‘S’-shaped radio emission centred on VLBI components (within the red rectangle). There also appear to be small-scale jets centred on the J1502P radio nucleus (red cross). Over-plotted are the JVL A 5 GHz contours (black) starting at 60 μ Jy per beam (4σ) and increasing in steps of 2σ . Also shown are the frame boundaries of the left panel. Negative contours are also shown at -4σ . The white ellipses in lower left of all three panels indicate the point spread function (PSF) full-width at half-maximum (FWHM) of the 5 GHz VLBI map. The grey ellipse in **a** shows the 1.7 GHz VLBI PSF, while the larger grey ellipses in **b** and **c** show the JVL A 5 GHz PSF.

$\approx 110 \text{ km s}^{-1}$, which is within the measured J1502S [O III] line width of 384 km s^{-1} . Given that the original J1502+1115 selection was based on double-peaked [O III] lines separated by 657 km s^{-1} , it is clear that neither the double-peaked [O III] nor the double-cored near-infrared imaging played any part in the selection of the much closer binary SMBH within J1502S. This implies that the only major selection criteria for the close-pair binary within J1502S are K-band magnitude ($K \approx 14$ – 16 mag), redshift ($z \approx 0.1$ – 0.4) and integrated 1.4 GHz flux density ($S_{1.4} \approx 1$ – 60 mJy). Based on simulations for radio continuum sources¹⁹, we expect a sky density of five such galaxies per square degree. Given that one sub-kpc binary was found from targeting six objects, this implies a sky density of $\Phi = 0.8^{+1.9}_{-0.7}$ per degree² for the given selection criteria, where we quote the formal Poisson 1σ uncertainties. Future VLBI surveys will be required to confirm this, given that the estimate is made from just six objects. While previous VLBI observations find a small fraction of binary/dual AGN⁹, they typically target bright sources which are usually associated with the most massive black holes ($M_{\text{BH}} \gtrsim 10^9 M_{\odot}$) for

which there is a low probability of a similarly massive companion black hole.

To date, there are very few confirmed binary/dual AGN, despite the expectations from our understanding of hierarchical structure formation through galaxy merging, combined with the observational evidence that every massive galaxy hosts a central SMBH²⁰. Of the binary/dual AGN with direct imaging confirmation, only three have sub-kpc projected separations and these systems are all at $z \lesssim 0.06$ (refs 10, 21, 22). In addition to these, there are a handful of candidate close-pair binaries based on light curve analyses, double-peaked broad lines and radio jet morphology^{23–25}, but these are not spatially resolved by current instruments. In Fig. 2 we show the projected spatial separations of the binary/dual AGN against redshift for a sample of sources that have been confirmed or discovered by spatially resolved X-ray, optical/infrared or radio imaging. While there may be a degree of subjectivity about which sources should appear in this list (some ambiguous cases have been omitted), it illustrates that the two VLBI-discovered systems have spatial

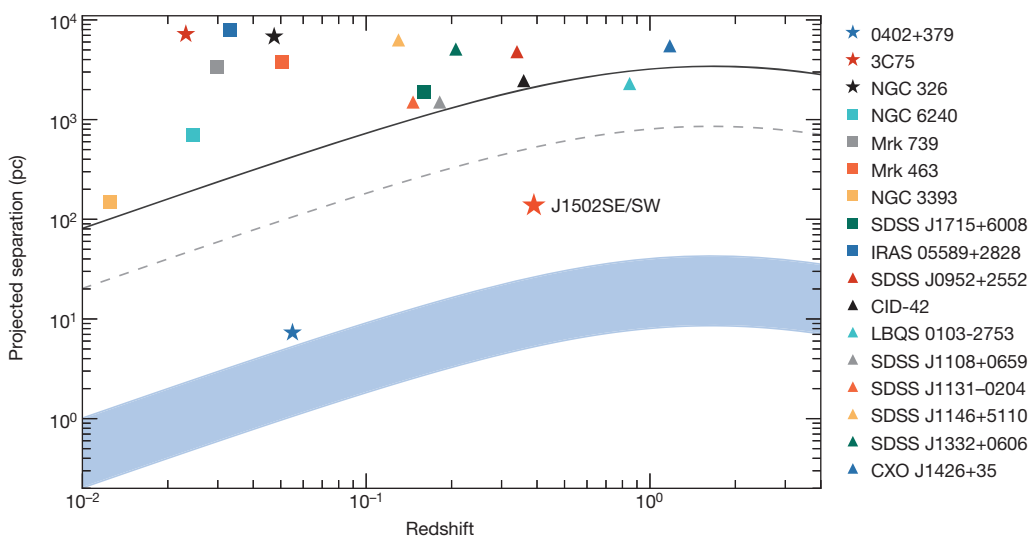


Figure 2 | Dual/binary AGN confirmed or discovered with direct imaging. Sample of binary/dual AGN confirmed or discovered by direct imaging at X-ray (squares), optical/infrared (triangles) and radio (stars) wavelengths. Typical spatial resolution limits of Chandra, the Hubble Space Telescope (HST) and VLBI are indicated by the solid and dashed lines, and shaded blue area, respectively. This illustrates that high-angular-resolution radio observations are able to survey a significantly larger cosmological volume for binary AGN with separations comparable to the black hole gravitational spheres of influence ($<10 \text{ pc}$ for $M_{\text{BH}} \approx 10^8 M_{\odot}$).

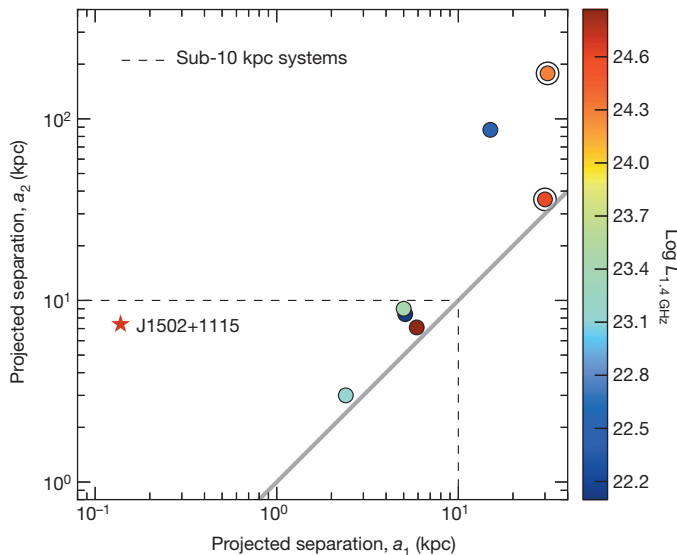


Figure 3 | Projected separations of candidate triple AGN. The axes denote the lowest (a_1) and second lowest (a_2) separations between AGN in the list of candidate triple AGN systems. The solid grey line indicates a ratio of unity. The dashed lines isolate the sub-10 kpc systems (see text). The colours correspond to the total integrated radio luminosity of the AGN plus the host galaxy (in units of W Hz^{-1}), since low-angular-resolution limits a fair AGN radio luminosity comparison. Nonetheless, the plot supports enhanced accretion within triple systems, particularly in the sub-10 kpc systems. J1502+1115 is an isolated region in the parameter space, probably owing to the bias against low separation for low-resolution optical/near-infrared and X-ray observations. The two black circled points (which are triple quasar systems) indicate upper radio luminosity limits from archival observations. The triple AGN candidates and their references are listed in Methods. Note that all sub-10 kpc systems are equidistant within a factor of 2, except for J1502+1115.

separations more than order of magnitude lower relative to X-ray and optical/infrared observations at similar redshifts. The J1502SE/SW binary has a period of $P_{\text{bin}} \approx 4$ Myr, assuming a circular orbit and major axis of 138 pc. This period is comparable to the lifetime of a typical radio source (~ 10 Myr).

Limited observational constraints exist on the abundance of triple black hole systems. There are only five triple AGN candidates with lowest (a_1) and second lowest (a_2) projected separations below 10 kpc ('sub-10 kpc systems' hereafter), which is the approximate effective radius for an elliptical galaxy. Figure 3 shows the lowest and second lowest projected separations for all the candidate triple AGN systems found in the literature (eight in total for separations $\lesssim 100$ kpc). Six of the eight known triple systems (and all the sub-10 kpc systems) have associated radio emission with 1.4 GHz luminosities in the range $\log L_{1.4} \approx 22$ –25 (here $L_{1.4}$ is in units of W Hz^{-1}). Like J1502+1115, some of these triple systems have radio emission associated with multiple nuclei. From the literature and archival data, we find that at least 9/15 ($>60\%$) of the sub-10 kpc SMBHs have associated radio emission. This is a lower limit for two reasons: (1) sensitivity of the available observations; and (2) some of the radio imaging available does not have sufficient angular resolution to resolve all three galaxies or nuclear cores in a triple system. The fraction of triple AGN with associated radio emission is significantly higher than the typical value of $\sim 10\%$ for AGN²⁶, suggesting that triple systems lead to higher accretion activity and consequently a higher chance of jet triggering. None of the sub-10 kpc systems were selected on the basis of their radio emission, implying no clear biasing selection effects (see details of host galaxies in Methods). Evidence for enhanced accretion has also been found for ~ 10 –100 kpc separation, hard-X-ray-selected dual AGN at low redshift²⁷. This is consistent with hydrodynamical simulations that find peak accretion and AGN luminosity occurring at the smallest separations (< 1 –10 kpc)²⁸. The above findings are therefore

consistent with the expected disruption to gas dynamics caused by multiple black hole orbits.

Multiple SMBH in-spiral and recoil rates, as well as the resulting orbital eccentricities, are predicted to dominate the stochastic gravitational wave background spectrum in the nHz– μHz range^{29,30}. Therefore, constraints on the number density of triple systems is important to refine the strategy of the future pulsar timing array experiments. This discovery demonstrates that high-angular-resolution radio observations have a unique advantage in searching for multiple black hole systems over large cosmological volumes, and therefore predicts that radio facilities like the Square Kilometre Array (as a stand-alone array or part of a VLBI network) will lead to a dramatic increase in the number of known low-separation ($\ll 1$ kpc) systems, as well as unresolved candidates for pointed gravitational wave experiments.

METHODS SUMMARY

J1502+1115 was observed at 1.7 and 5 GHz with the European VLBI Network. The 305 m Arecibo observatory was included in the 1.7 GHz observation. The 5 GHz observation used the e-VLBI technique, which performs real-time correlation. The correlation averaging time was 2 s and 4 s for the 5 GHz and 1.7 GHz visibilities respectively and we used 32 delay steps (lags) per sub-band. There were 4×8 MHz subbands observed in both left and right circular polarizations with two-bit sampling. Both observations used 4C39.25 as a fringe-finding source and the nearby (angular distance $\Delta\theta \approx 51$ arcmin) source J1504+1029 as a phase calibrator. Both maps have a brightness scale uncertainty of $\sim 10\%$. The VLBI observations at 1.7 and 5 GHz were taken 18 months apart (and 23 months for the Arecibo run). The 5 GHz JVLA observations were performed in A-configuration and all details thereof are described in the literature¹⁵. We re-calibrated these data using the CASA software package and achieved flux densities consistent with previously published values¹⁵, within the uncertainties. The residual map shown in Fig. 1c was generated by taking the median of 500 realizations of maps made by subtracting point sources directly from the JVLA 5 GHz visibilities. The point source positions and flux densities were derived from the VLBI 5 GHz components and the JVLA 5 GHz map for the J1502S and J1502P components, respectively. Each realization added a random offset in flux density and position, defined by their 1σ uncertainties and drawn from a Gaussian distribution, so that the effect of calibration and deconvolution errors on the residual structure could be constrained. The 5 GHz VLBI observation was made 11 weeks before the JVLA 5 GHz observation¹⁵, limiting the uncertainty due to source variability in the comparison of the two.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 February; accepted 1 May 2014.

Published online 25 June 2014.

1. Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).
2. Volonteri, M., Haardt, F. & Madau, P. The assembly and merging history of supermassive black holes in hierarchical models of galaxy formation. *Astrophys. J.* **582**, 559–573 (2003).
3. Begelman, M. C., Blandford, R. D. & Rees, M. J. Massive black hole binaries in active galactic nuclei. *Nature* **287**, 307–309 (1980).
4. Kulkarni, G. & Loeb, A. Formation of galactic nuclei with multiple supermassive black holes at high redshifts. *Mon. Not. R. Astron. Soc.* **422**, 1306–1323 (2012).
5. Tonry, J. L. Constraints on the orbits of multiple nuclei in brightest cluster galaxies. *Astrophys. J.* **279**, 13–18 (1984).
6. Barth, A. J., Bentz, M. C., Greene, J. E. & Ho, L. C. An offset Seyfert 2 nucleus in the minor merger system NGC 3341. *Astrophys. J.* **683**, L119–L122 (2008).
7. Liu, X., Shen, Y. & Strauss, M. A. Cosmic train wreck by massive black holes: discovery of a kiloparsec-scale triple active galactic nucleus. *Astrophys. J.* **736**, L7 (2011).
8. Schawinski, K. *et al.* Evidence for three accreting black holes in a galaxy at $z \sim 1.35$: a snapshot of recently formed black hole seeds? *Astrophys. J.* **743**, L37 (2011).
9. Burke-Spolaor, S. A radio census of binary supermassive black holes. *Mon. Not. R. Astron. Soc.* **410**, 2113–2122 (2011).
10. Rodríguez, C. *et al.* A compact supermassive binary black hole system. *Astrophys. J.* **646**, 49–60 (2006).
11. Wyithe, J. S. B. & Loeb, A. Low-frequency gravitational waves from massive black hole binaries: predictions for LISA and pulsar timing arrays. *Astrophys. J.* **590**, 691–706 (2003).
12. Sesana, A. Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band. *Mon. Not. R. Astron. Soc.* **433**, L1–L5 (2013).
13. Smith, K. L. *et al.* A search for binary active galactic nuclei: double-peaked [O III] AGNs in the Sloan Digital Sky Survey. *Astrophys. J.* **716**, 866–877 (2010).

14. Fu, H., Myers, A. D., Djorgovski, S. G. & Yan, L. Mergers in double-peaked [O III] active galactic nuclei. *Astrophys. J.* **733**, 103–109 (2011).
15. Fu, H. *et al.* A kiloparsec-scale binary active galactic nucleus confirmed by the expanded Very Large Array. *Astrophys. J.* **740**, L44 (2011).
16. Blandford, R. D. & Königl, A. Relativistic jets as compact radio sources. *Astrophys. J.* **232**, 34–48 (1979).
17. Hjellming, R. M. & Johnston, K. J. An analysis of the proper motions of SS 433 radio jets. *Astrophys. J.* **246**, L141–L145 (1981).
18. Häring, N. & Rix, H.-W. On the black hole mass–bulge mass relation. *Astrophys. J.* **604**, L89–L92 (2004).
19. Wilman, R. J. *et al.* A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes. *Mon. Not. R. Astron. Soc.* **388**, 1335–1348 (2008).
20. Kormendy, J. & Richstone, D. Inward bound—the search for supermassive black holes in galactic nuclei. *Annu. Rev. Astron. Astrophys.* **33**, 581–624 (1995).
21. Komossa, S. *et al.* Discovery of a binary active galactic nucleus in the ultraluminous infrared galaxy NGC 6240 using Chandra. *Astrophys. J.* **582**, L15–L19 (2003).
22. Fabbiano, G., Wang, J., Elvis, M. & Risaliti, G. A close nuclear black-hole pair in the spiral galaxy NGC 3393. *Nature* **477**, 431–434 (2011).
23. Kaastra, J. S. & Roos, N. Massive binary black-holes and wiggling jets. *Astron. Astrophys.* **254**, 96–98 (1992).
24. Valtonen, M. J. *et al.* A massive binary black-hole system in OJ287 and a test of general relativity. *Nature* **452**, 851–853 (2008).
25. Boroson, T. A. & Lauer, T. R. A candidate sub-parsec supermassive binary black hole system. *Nature* **458**, 53–55 (2009).
26. Jiang, L. *et al.* The radio-loud fraction of quasars is a strong function of redshift and optical luminosity. *Astrophys. J.* **656**, 680–690 (2007).
27. Koss, M. *et al.* Understanding dual active galactic nucleus activation in the nearby Universe. *Astrophys. J.* **746**, L22 (2012).
28. Van Wassenhove, S. *et al.* Observability of dual active galactic nuclei in merging galaxies. *Astrophys. J.* **748**, L7 (2012).
29. Hoffman, L. & Loeb, A. Dynamics of triple black hole systems in hierarchically merging massive galaxies. *Mon. Not. R. Astron. Soc.* **377**, 957–976 (2007).
30. Blecha, L., Cox, T. J., Loeb, A. & Hernquist, L. Recoiling black holes in merging galaxies: relationship to active galactic nucleus lifetimes, starbursts and the $M_{\text{BH}}-\sigma_*$ relation. *Mon. Not. R. Astron. Soc.* **412**, 2154–2182 (2011).

Acknowledgements We thank J. Magorrian, A. Karastergiou, S. Ransom and B. Fanaroff for discussions. The European VLBI Network is a joint facility of European, Chinese, South African and other radio astronomy institutes funded by their national research councils. e-VLBI research infrastructure in Europe was supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number RI-261525 NEXPReS. The financial assistance of the South African SKA Project (SKA SA) towards this research is acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the SKA SA. Z.P. and S.F. acknowledge funding from the Hungarian Scientific Research Fund (OTKA 104539). Z.P. is grateful for funding support from the International Space Science Institute.

Author Contributions R.P.D. was the Principal Investigator of the project and wrote the paper. Z.P. helped design, schedule, observe and calibrate the EVN observations as well as interpret the results. M.J.J. helped interpret the cosmological and astrophysical importance of the discovery and contributed significantly to the text. M.C. and R.P.F. helped in the binary SMBH interpretation and the physics thereof. S.F. calibrated the 1.7 GHz EVN observation and helped in the VLBI component interpretation. G.B. and I.H. helped with the technical aspects of the JVLA re-analysis and the subtle interferometric effects at play. H.R.K. helped in the VLBI and GMRT proposals and wrote the software to calibrate the GMRT observation. K.G. and C.R. observed and calibrated the 16 GHz AMI observations. All authors contributed to the analysis and text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.P.D. (roger.deane@ast.uct.ac.za).

METHODS

Observations using e-VLBI 5 GHz. J1502+1115 was observed at 5 GHz with the European VLBI Network using the e-VLBI technique on 12 April 2011 for approximately 1.5 h. The participating telescopes were Effelsberg (Germany), Jodrell Bank (Mk II, UK), Hartebeesthoek (South Africa), Sheshan (China), Medicina (Italy), Onsala (Sweden), Toruń (Poland), Yebes (Spain) and the phased-array Westerbork Synthesis Radio Telescope (WSRT, Netherlands). The data were streamed real-time to the EVN Data Processor at the Joint Institute for VLBI in Europe (JIVE) at a 1,024 Mbps data rate. The correlation averaging time was 2 s and we used 32 delay steps (lags) per sub-band. The coordinates used for correlation were the J1502+1115 1.4 GHz centroid determined from the Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) survey³¹. J1502SE/SW are detected sufficiently near the phase centre ($\Delta\theta \lesssim 1.3$ arcsec) that time and bandwidth smearing losses are negligible. There were 4×8 MHz sub-bands observed in both LCP and RCP polarizations with 2-bit sampling. The initial clock searching was carried out before the experiment using the fringe-finder source 4C39.25. J1502+1115 was phase-referenced to the nearby ($\Delta\theta \approx 51$ arcmin) calibrator J1504+1029 for 1.3 min per 5 min calibrator-target cycle. The J2000 coordinates of the reference source (RA = 15 h 04 min 24.979782 s, dec. = $+10^\circ 29' 39.19858''$) were taken from the Very Long Baseline Array (VLBA) Calibrator Survey (<http://www.vlba.nrao.edu/astro/calib/index.shtml>) and have a positional uncertainty of ~ 0.02 mas. The resultant map has a Briggs-weighted (robust = 0) synthesized beam with a 9.1×6.1 mas² FWHM at a position angle of 76° east of north. The r.m.s. noise in the Briggs-weighted (robust = 0) 5 GHz map is ~ 36 μ Jy per beam and the brightness scale uncertainty is $\sim 10\%$. The J1502P component is not detected in the VLBI observations with a 5σ upper limit of $S_{5\text{GHz}} \lesssim 180$ μ Jy per beam. Components J1502SE and J1502SW are detected with flux densities of $S_{5\text{GHz}} = 857 \pm 49$ and 872 ± 49 μ Jy respectively, where the uncertainties are the quadrature sum of the fitting uncertainty and the r.m.s. of the map measured ~ 2 arcsec from the phase centre. This e-VLBI observation was made 11 weeks before the JVL A 5 GHz observation¹⁵, limiting the uncertainty due to source variability in the comparison of the two.

Observations using VLBI 1.7 GHz. Based on the detection of two compact components in the e-VLBI 5 GHz map, we observed J1502S at 1.7 GHz using the European VLBI Network. The observations were performed on 24 October 2012 and 13 March 2013. Antennas included in the VLBI array for the 2012 run were the same as the 5 GHz e-VLBI observations, as well as Urumqi (China) and the Russian KVAZARNetwork stations Svetloe, Zelenchukskaya and Badary. The second epoch observations included the following antennas: Arecibo, Effelsberg, Jodrell Bank, Onsala, Medicina, Toruń, Noto and WSRT. J1502S was observed for approximately 3 h in total. The data were 2-bit sampled. The data in this case were recorded with Mark5As and correlated at a later time at JIVE. There were 128 delay steps per sub-band and the integration time was 4 s. The same fringe-finder, phase reference target and duty cycle was used as in the e-VLBI observation. The coordinates used for correlation were determined from the e-VLBI 5 GHz detection. The 1.7 and 5 GHz VLBI centroids are consistent within the fitting uncertainties ($\lesssim 1.5$ mas, 7.5 pc) for non-self-calibrated 1.7 GHz *uv*-data. Including the Arecibo baselines in the 1.7 GHz *uv*-data and self-calibrating all Arecibo baselines leads to a higher angular resolution and fidelity map, but with compromised absolute astrometry. The resultant offset is ($\Delta\text{RA}, \Delta\text{dec.}$) = (0.55 mas, 3.15 mas). This shift is corrected for in the map shown of J1502SE/SW in the left panel of Fig. 1 of the main paper. The measured flux densities for self-calibrated and un-self-calibrated data are consistent within the fitting uncertainties. The naturally weighted map has an r.m.s. of ~ 31 μ Jy per beam and a synthesized beam FWHM of 6.5×3.9 mas² at a position angle of $\sim 42^\circ$ east of north. The J1502P component is not detected in the VLBI observations with a 5σ upper limit of $S_{1.7\text{GHz}} \lesssim 170$ μ Jy per beam. Components J1502SE and J1502SW are detected with flux densities of $S_{1.7\text{GHz}} = 954 \pm 50$ and 920 ± 49 μ Jy respectively, where the uncertainties are the quadrature sum of the fitting uncertainty and the r.m.s. of the map measured ~ 2 arcsec from the phase centre. The VLBI observations at 1.7 and 5 GHz were taken 18 months apart (and 23 months for the Arecibo run), yielding an apparent proper motion upper limit of $\Delta\mu \lesssim 5$ μ pc yr⁻¹.

Observations using Arcminute Microkelvin Imager (AMI) 16 GHz. J1502+1115 was observed on 6 December 2012 for 12 h by the AMI Large Array³² for approximately 1.5 h. The telescope observes between 13.5 and 18 GHz in six spectral channels of 0.75 GHz bandwidth. The source J1504+1029 was observed for 100 s every 600 s for accurate phase calibration. Flux density calibration was performed using short observations of 3C 286 which were calibrated against VLA measurements³³. The data were flagged and calibrated using REDUCE, a local software package developed for AMI. The calibrated data were then mapped and CLEANED in AIPS using the IMAGR task. The integrated flux density of the detection was $S_{\text{int}} \approx 1.17$ mJy. The final thermal noise on the continuum map was 35 μ Jy per beam and a flux scale uncertainty of 5%. The map has not been corrected for the telescope primary beam response, which is well modelled by a Gaussian of FWHM 5.6 arcmin. The CLEAN restoring beam has a FWHM of 64×23 arcsec at a position angle of 9° . We do not

expect confusion with other sources within this beam based on the higher resolution and sensitivity JVL A continuum 8 GHz map.

Observations using Giant Metrewave Radio Telescope (GMRT) 610 MHz. J1502+1115 was observed with the GMRT array on 17 December 2012 at 610 MHz for approximately 45 min. J1502+1115 was phase referenced to the nearby ($\Delta\theta \approx 51$ arcmin) calibrator J1504+1029. Amplitude calibration was performed using 3C 286. Data processing was performed with an automated calibration and imaging pipeline based on PYTHON, AIPS and PARSELTONGUE³⁴ which has been specially developed to process GMRT data. The full technical description of this calibration and imaging pipeline is extensive and is presented in detail elsewhere³⁵. The integrated flux density of J1502+1115 at 610 MHz is $S_{\text{int}} = 22.84 \pm 0.95$ mJy, where the uncertainty is the quadrature sum of the fitting uncertainty and the local noise near the J1502+1115 detection. The CLEAN restoring beam is 5.5×4.4 arcsec at a position angle of 59° , and therefore does not spatially resolve J1502P and J1502S.

Observations using JVL A. The JVL A observations were performed on 29 June 2011 at 1.4 and 5 GHz. The 8 GHz observations were performed on 14 June 2011. All three of these bands were observed while the JVL A was in A-configuration. The details of the observations are reported in the literature¹⁵. We re-calibrated all three observations using the CASA software package and achieved flux densities consistent with previously determined values¹⁵, within the uncertainties.

Radio spectra of radio components. In Extended Data Fig. 1 we show the radio spectra of the various components. The component J1502P has a steep spectrum and is unlikely to contribute to the measured 15.7 GHz AMI flux density. It therefore appears that the AMI map is dominated by the flat-spectrum cores.

JVL A 5 GHz residual map. The residual map shown in the bottom right panel of Fig. 1 was generated by taking the median of 500 realizations of maps made by subtracting point sources directly from the JVL A 5 GHz visibilities. The point source positions and flux densities were derived from the VLBI 5 GHz components (observed 11 weeks before the JVL A 5 GHz observation) and the Briggs-weighted (robust = 0) 5 GHz map of the JVL A 5 GHz data itself for the J1502S and J1502P components respectively. Each realization added a random offset in flux density and position, defined by their 1σ uncertainties and drawn from a Gaussian distribution, so that the effect of calibration and deconvolution errors on the residual structure could be constrained. Note the JVL A 8 GHz residual map shows similar structure, but at a lower signal-to-noise ratio. This analysis was performed using the MeqTrees software package³⁶ (www.meqtrees.net). The JVL A-detected jet emission is not detected in the VLBI maps, due to insufficient brightness temperature sensitivity.

JVL A 5 GHz inner jet. The JVL A residual map discussed above (and shown in Fig. 1c) was generated using Briggs weighting (robust = 1). In Extended Data Fig. 3, we show the same JVL A 5 GHz residuals but with Briggs weighting (robust = 0). This results in higher angular resolution (synthesized beam FWHM = 0.39×0.32 arcsec², position angle = 45° east of north), but with poorer sensitivity ($1\sigma \approx 30$ μ Jy per beam), or more importantly, poorer brightness temperature sensitivity. This results in a map that traces the higher brightness temperature emission which we use to probe the central region of J1502S (see Extended Data Fig. 3). This figure shows the JVL A 5 GHz inner jet at a position angle of $\sim 45^\circ$ east of north. This is significantly misaligned with the position angle of the vector between J1502SE and J1502SW ($\theta_{\text{J1502SE/SW}} \approx 87^\circ$). J1502SE is within ~ 20 mas of the brightest pixel in the JVL A 5 GHz residual map, well within the absolute JVL A astrometry of 100 mas³⁷. This supports the view that the 'S'-shaped jet emission seen in Fig. 1c stems from one of the two VLBI-detected cores. The large misalignment is inconsistent with any scenario where the two VLBI components are interpreted as young hotspots stemming from a single, undetected core. Furthermore, the straight northeast inner jet with projected length ~ 0.4 arcsec (~ 2.1 kpc) and a jet axis position angle of $\sim 45^\circ$ is also inconsistent with a scenario where one of the VLBI components is an AGN core and the other a hotspot along the jet trajectory. A detailed discussion is also included in the following section.

Alternative interpretations of the parsec-scale radio emission in J1502S. In the two sub-sections below we distinguish between jets and hotspots to simplify comparison with the literature.

First, double-jet or core-jet scenarios. A possible interpretation of the double flat-spectrum structure of J1502S is that we observe the synchrotron self-absorbed bases of a pair of oppositely directed jets from a single AGN. Similar structures have been seen in nearby type-II AGN, as well as in the Galactic X-ray binary system SS 433³⁸. We may exclude this scenario for a number of reasons. (1) We do not see elongated core-jet structure in either of the two components, which is contrary to what is typical of self-absorbed AGN cores. (2) We would not expect these to be unresolved at VLBI scales given the fact that the two components are separated by 26 mas, and the separation between the two jet sides should be comparable to the size of the core itself⁶. (3) The 138 pc physical separation between the two components is an order of magnitude larger than the typical size of radio cores (~ 1 –10 pc) seen in the brightest AGN^{39,40}.

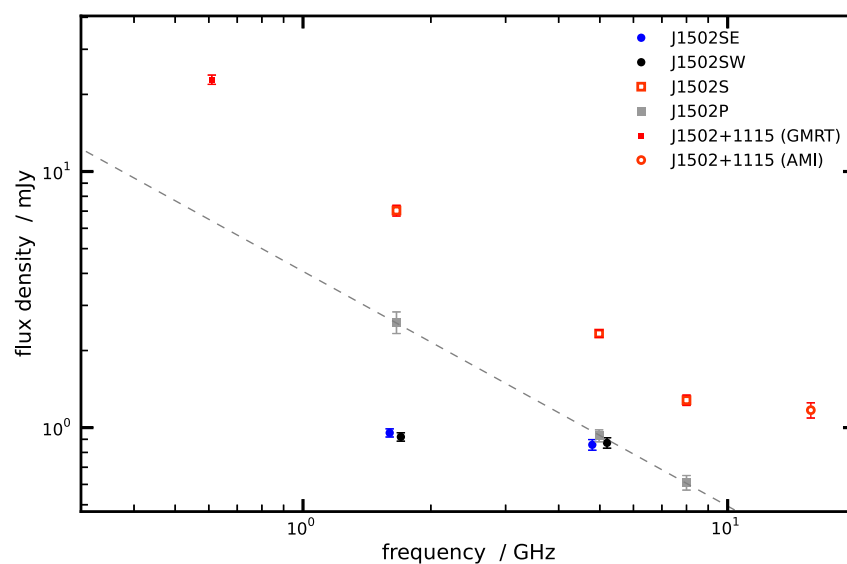
Second, double-hotspot or core-hotspot scenarios. There are a number of arguments against the scenario where the two compact J1502SE/SW components are hotspots as observed in the class of compact symmetric objects (CSOs); or alternatively that one component is a hotspot and the other the core. We outline the most important points. (1) Both components have flat spectra, which is uncharacteristic of hotspots which have typical spectral indices of $\alpha_{\text{hotspots}} < -0.5$ ⁴¹. Spectral indices are even steeper for CSOs with a projected linear size (hotspot separation) greater than $\sim 100 \text{ pc}$ ⁴¹. (2) The population of CSOs have typical luminosities of $10^{27} \text{ W Hz}^{-1}$, which is 4 orders of magnitude larger than J1502SE/SW^{41,42}. While this may be a selection effect, the lowest luminosity CSO that we have found in the literature has a 1.4 GHz radio luminosity that is 1 order of magnitude larger than that of J1502SE/SW⁴¹. (3) Similar to the double-jet/core-jet scenario discussed in the previous section, higher frequency observations probe nearer the base of the jet since it becomes optically thin closer to the central black hole^{39,40}. In the core-hotspot scenario where the distance is as large as 138 pc, this would result in an observable frequency dependent core-shift (that is, decreasing separation between the two components) which we do not see: the separation of J1502SE and J1502SW appears independent of observing frequency. (4) In the double-hotspot scenario, there is no core emission detected (to $L_{1.7\text{GHz}} \approx 10^{22} \text{ W Hz}^{-1}$). (5) The straight $\sim 2\text{-kpc}$ -long inner jet of J1502S has a 45° misalignment with the vector between J1502SE/SW. This would imply a dramatic, almost instantaneous change in angular momentum. (6) Assuming that one or both J1502SE/SW are young hotspots; this would imply that we happened to observe J1502+1115 within a very short time period since this dramatic change in angular momentum. If we assume that the jet axis is in the plane of the sky and a very conservative jet speed of $0.01\text{--}0.1c$, the shift in momentum must have occurred within the last $(2\text{--}20) \times 10^3 \text{ yr}$. For a nominal radio source lifetime of $\sim 10^7 \text{ yr}$, this represents a $0.02\text{--}0.2\%$ probability. Indeed, hotspot velocities from young radio sources are typically $0.3\text{--}0.5c$, and almost never below $0.1c$, corroborating our claim that this is a very conservative estimate⁴¹.

So in summary, in order for a core-hotspot or double-hotspot alternative to be true, the hotspot(s) would not only need to have highly uncharacteristic flat spectral indices (and potentially an order of magnitude larger core size compared to what is observed in the highest luminosity radio sources), but also to have been observed almost instantaneously after a major shift in angular momentum has occurred. Although the binary SMBH interpretation may be considered a (hitherto) exotic one, combining the above arguments with rotationally-symmetric 'S'-shaped flux strongly favours the binary SMBH interpretation.

Triple AGN candidates. The triple AGN candidates plotted in Fig. 3 of the main text and the relevant references are SDSS J1027+1749⁷, NGC 3341^{6,43}, QQ 1429-008⁴⁴,

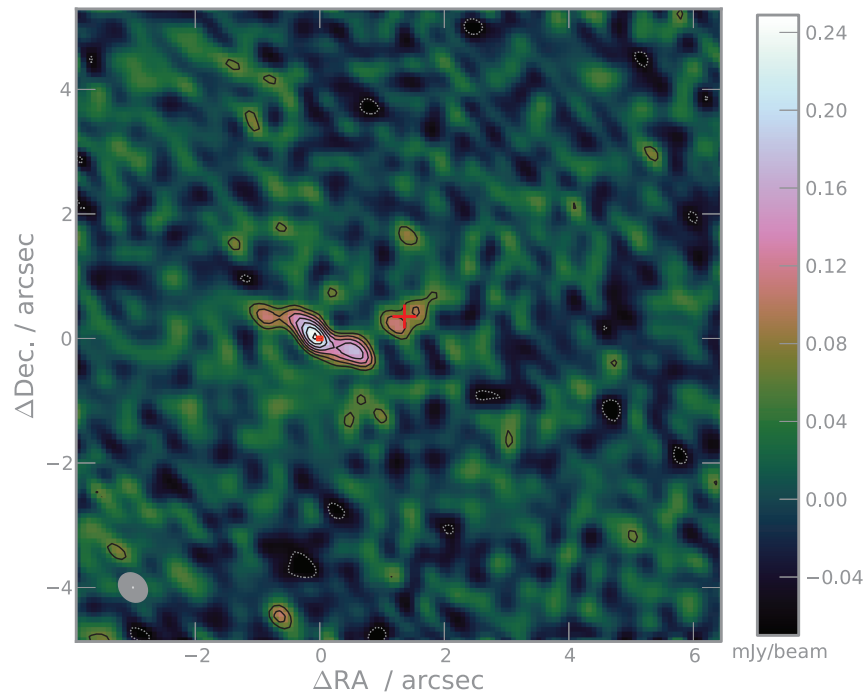
GOODS J123652.77+621354.7⁸, QQ J1519+0627⁴⁵, NGC 6166^{5,46} and the NGC 835 group²⁷. The 1σ upper limits on 1.4 GHz radio luminosity for the two triple quasars are determined from the local noise estimate around their respective sky coordinates in VLA FIRST survey³¹. There is a broad range in redshifts for this list of triple AGN candidates spanning from $z \approx 0.005$ to ~ 2 . Primary host galaxy morphology does not appear to play a strong role in preferentially selecting the six triple AGN systems with radio emission (3 disks, 2 ellipticals, 1 cD).

31. Becker, R. H., White, R. L. & Helfand, D. J. The FIRST Survey: faint images of the radio sky at twenty centimeters. *Astrophys. J.* **450**, 559–577 (1995).
32. Zwart, J. T. L. *et al.* The Arcminute Microkelvin Imager. *Mon. Not. R. Astron. Soc.* **391**, 1545–1558 (2008).
33. Perley, R. A. & Butler, B. J. An accurate flux density scale from 1 to 50 GHz. *Astrophys. J.* **204** (Supp.), 19 (2013).
34. Kettenis, M., van Langevelde, H. J., Reynolds, C. & Cotton, B. in *Astronomical Data Analysis Software and Systems XV* (eds Gabriel, C., Arviset, C., Ponz, D. & Enrique, S.) 497–500 (ASP Conf. Ser. 351, Astronomical Society of the Pacific, 2006).
35. Mauch, T. *et al.* A 325-MHz GMRT survey of the Herschel-ATLAS/GAMA fields. *Mon. Not. R. Astron. Soc.* **435**, 650–662 (2013).
36. Noordam, J. E. & Smirnov, O. M. The MeqTrees software system and its use for third-generation calibration of radio interferometers. *Astron. Astrophys.* **524**, A61 (2010).
37. Ulvestad, J. S., Perley, R. A. & Chandler, C. J. VLA Observational Status Summary. <https://science.nrao.edu/facilities/vla/docs/manuals/oss2013b/performance/positional-accuracy> (2009).
38. Paragi, Z. *et al.* The inner radio jet region and the complex environment of SS433. *Astron. Astrophys.* **348**, 910–916 (1999).
39. Lobanov, A. P. Ultracompact jets in active galactic nuclei. *Astron. Astrophys.* **330**, 79–89 (1998).
40. Hada, K. *et al.* An origin of the radio jet in M87 at the location of the central black hole. *Nature* **477**, 185–187 (2011).
41. An, T. & Baan, W. A. The dynamic evolution of young extragalactic radio sources. *Astrophys. J.* **760**, 77 (2012).
42. O'Dea, C. P. The compact steep-spectrum and gigahertz peaked-spectrum radio sources. *Publ. Astron. Soc. Pacif.* **110**, 493–532 (1998).
43. Bianchi, S. *et al.* The NGC 3341 minor merger: a panchromatic view of the active galactic nucleus in a dwarf companion. *Mon. Not. R. Astron. Soc.* **435**, 2335–2344 (2013).
44. Djorgovski, S. G. *et al.* Discovery of a probable physical triple quasar. *Astrophys. J.* **662**, L1–L5 (2007).
45. Farina, E. P., Montuori, C., Decarli, R. & Fumagalli, M. Caught in the act: discovery of a physical quasar triplet. *Mon. Not. R. Astron. Soc.* **431**, 1019–1025 (2013).
46. Ge, J. & Owen, F. N. Faraday rotation in cooling flow clusters of galaxies. 2: Survey. *Astron. J.* **108**, 1523–1533 (1994).



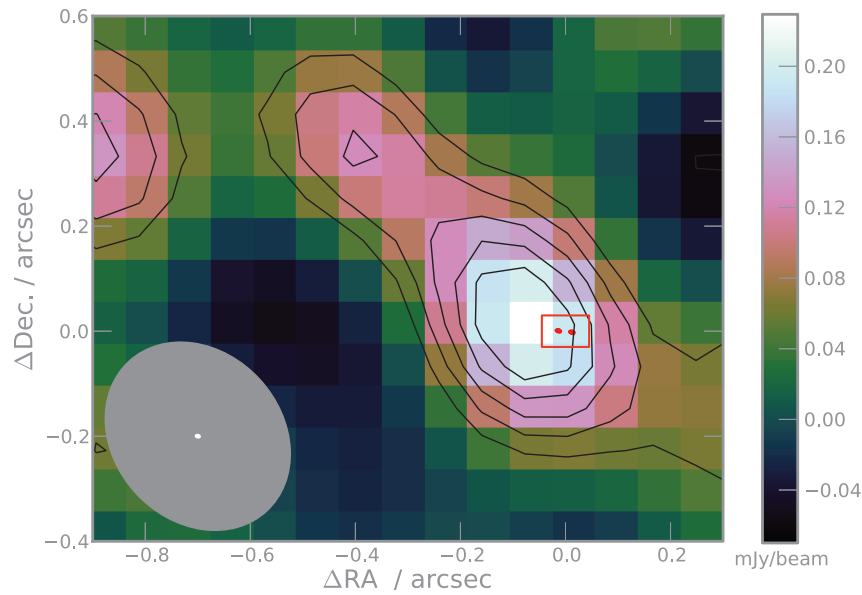
Extended Data Figure 1 | Radio spectrum of the radio components in J1502+1115. Both J1502S (open red squares) and J1502P (grey squares) are steep-spectrum radio sources between 1.4 and 8 GHz, as measured by JVL A observations. The two flat spectrum cores (J1502SE/SW, filled circles) are the likely cause of the flattened spectrum of J1502S at higher frequency, as

measured by the AMI 15.7 GHz detection labelled as J1502+1115 (AMI). Error bars, ± 1 s.d. on the flux measurement. J1502+1115 (GMRT) indicates the 610 MHz detection using the GMRT (Methods). Note that J1502SE and J1502SW are shown marginally offset in frequency purely for clearer illustration.



Extended Data Figure 2 | Larger field-of-view JVL 5 GHz map of J1502+1115 to demonstrate map fidelity. The colour scale shows the same JVL 5 GHz residuals shown in Fig. 1c but with the full 128×128 pixel median map generated from the Monte Carlo realizations. The small filled red square indicates the map boundary of the VLBI map shown in Fig. 1a. The red

cross denotes the centroid of the point source subtracted J1502P component. Contour levels are the same as in Fig. 1c. The grey ellipse (lower left) represents the FWHM of the Briggs-weighted (robust = 1) PSF, while the white dot shows the VLBI 5 GHz PSF.



Extended Data Figure 3 | Zoom-in of the high-brightness-temperature inner jet emission of J1502S. The colour scale shows the same JVL A 5 GHz residuals shown in Fig. 1c but imaged with Briggs uv -weighting (robust = 0) to highlight the position angle of the inner northeast J1502S jet. This is misaligned with the vector between J1502SE and J1502SW (red dots) by $\sim 45^\circ$.

The black JVL A 5 GHz contours start at $60 \mu\text{Jy}$ per beam ($\sim 2\sigma$) and increase in steps of 1σ . The grey ellipse (lower left) represents the FWHM of the Briggs-weighted (robust = 0) PSF, while the white ellipse shows the VLBI 5 GHz PSF. The red square indicates the map boundary of the VLBI map shown in Fig. 1a.

Normal-state nodal electronic structure in underdoped high- T_c copper oxides

Suchitra E. Sebastian¹, N. Harrison², F. F. Balakirev², M. M. Altarawneh^{2,3}, P. A. Goddard⁴, Ruixing Liang^{5,6}, D. A. Bonn^{5,6}, W. N. Hardy^{5,6} & G. G. Lonzarich¹

An outstanding problem in the field of high-transition-temperature (high- T_c) superconductivity is the identification of the normal state out of which superconductivity emerges in the mysterious underdoped regime¹. The normal state uncomplicated by thermal fluctuations can be studied using applied magnetic fields that are sufficiently strong to suppress long-range superconductivity at low temperatures^{2,3}. Proposals in which the normal ground state is characterized by small Fermi surface pockets that exist in the absence of symmetry breaking^{1,4–8} have been superseded by models based on the existence of a superlattice that breaks the translational symmetry of the underlying lattice^{7–15}. Recently, a charge superlattice model that positions a small electron-like Fermi pocket in the vicinity of the nodes (where the superconducting gap is minimum)^{8,9,16,17} has been proposed as a replacement for the prevalent superlattice models^{10–14} that position the Fermi pocket in the vicinity of the pseudogap at the antinodes (where the superconducting gap is maximum)¹⁸. Although some ingredients of symmetry breaking have been recently revealed by crystallographic studies, their relevance to the electronic structure remains unresolved^{19–21}. Here we report angle-resolved quantum oscillation measurements in the underdoped copper oxide $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. These measurements reveal a normal ground state comprising electron-like Fermi surface pockets located in the vicinity of the nodes, and also point to an underlying superlattice structure of low frequency and long wavelength with features in common with the charge order identified recently by complementary spectroscopic techniques^{14,19–22}.

The normal ground-state electronic structure revealed by our measurements is summarized in Fig. 1. Quantum oscillations measured in the contactless electrical resistivity of the underdoped copper oxide $\text{YBa}_2\text{Cu}_3\text{O}_{6.56}$ are shown in Fig. 2a. The prominent oscillatory beat structure reveals a sizeable frequency spread^{8,16}. Measurements are made as a function of magnetic field up to 85 T over a wide range of field orientations as defined by the polar angle θ and the azimuthal angle ϕ . This data may be used to identify characteristics of the momentum-space electronic structure in the way that, for example, X-ray data may be used to identify the real-space lattice structure of a crystal. In particular, it can be employed to identify the correct Fermi surface model²³. In the case of underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ which has a primitive orthorhombic lattice, Fig. 2b illustrates the quasi-two-dimensional electronic structure characterized by a cylindrical Fermi surface with fundamental neck and belly warping that would be expected in the absence of a superlattice (Fig. 3a–d). A distinguishing characteristic of this neck and belly Fermi surface is that it gives rise to a resonance in the amplitude of quantum oscillations at a polar angle known as the Yamaji angle²³. The resonance in amplitude is expected to grow with the size of the frequency spread, and its location is expected at approximately 60° in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ given the measured mean diameter of the cylinder (corresponding to the dominant quantum oscillation frequency) and the c -axis lattice constant (Fig. 2b, c). Strikingly, however, despite the sizeable frequency spread of the quantum oscillation spectrum (Fig. 2a), higher angle data in Fig. 2a reveal

the Yamaji resonance anticipated for a fundamental neck and belly Fermi surface geometry to be absent, pointing to an alternative Fermi surface geometry. Supporting analyses are shown in the Methods and in Extended Data Figs 1, 2, 3 and 4.

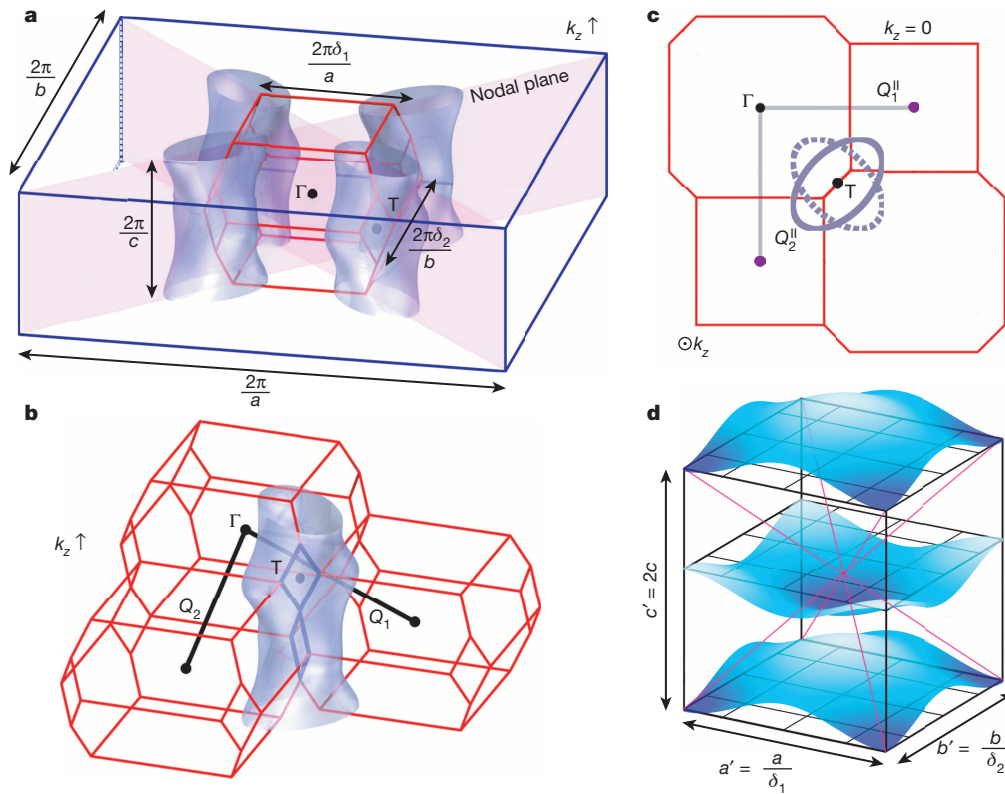
To arrive at the correct Fermi surface model we are guided by the following key experimental findings: (1) as shown in Fig. 4a, the replacement of the expected Yamaji resonance indeed by an anti-resonance; (2) as shown in Fig. 4b, the fourfold anisotropy in ϕ dependence of the quantum oscillation amplitude for different values of θ , with a maximization of quantum oscillation amplitude along the a and b crystallographic directions and minimization along the diagonal directions, for example, $\phi = 45^\circ$; and (3) as shown in Fig. 4b, the enhancement in fourfold anisotropy as a function of azimuthal angle with increasing values of polar angle θ .

While the observed experimental features are in sharp variance with those expected for a dominant fundamental neck and belly Fermi surface geometry, we find that a dominant staggered twofold Fermi surface with warping maxima along the diagonal directions in momentum space (Fig. 1a–c, Fig. 3f–h) predicts the experimental features that we have observed. The simulation of the quantum oscillation waveform (Fig. 3h and 4 and Methods) shows that such a staggered twofold Fermi surface geometry yields first an anti-resonance in the quantum oscillation amplitude at a special value of θ along the diagonal directions in ϕ , second, a fourfold anisotropy in the quantum oscillation amplitude as a function of ϕ that is minimized along the diagonal directions, and finally an enhancement in the fourfold anisotropy as a function of ϕ with increased values of θ . Layered materials with examples of both fundamental neck and belly and staggered twofold Fermi surface geometries are discussed in the Methods.

Simulations of a staggered twofold Fermi surface geometry with modulation amplitude $\Delta F_{\text{twofold}} \approx 15$ T agree well with the measured quantum oscillation data as a function of all three parameters B , θ and ϕ (Fig. 4, Extended Data Table 2). In addition to the amplitude damping factor for a staggered twofold geometry ($R_{\text{w}}^{\text{twofold}}$), conventional damping factors are included that are used to describe layered materials, which arise from thermal smearing, impurity scattering or quenched inhomogeneities, Zeeman splitting and magnetic breakdown. The quantum oscillatory frequency spread is captured by Fermi surface splitting from a finite bilayer or spin orbit coupling. More details of the simulation are given in the Methods, Extended Data Figs 2c, 4, 5, 6 and 7 and Extended Data Tables 1 and 2.

The staggered twofold Fermi surface geometry we observe would not be expected to dominate within the primitive orthorhombic unit cell of underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. Yet just such a geometry would arise from the unique symmetry of the corner T point in the Brillouin zone of a body-centred orthorhombic unit cell defined in Figs 1 and 3e–h (see Methods). At this special corner T point, the twofold in-plane rotational symmetry alternates by 90° between adjacent symmetry planes. The diagonal orientation of the maximal warping directions of the Fermi surface (shown

¹Cavendish Laboratory, Cambridge University, JJ Thomson Avenue, Cambridge CB3 0HE, UK. ²National High Magnetic Field Laboratory, Los Alamos National Laboratory (LANL), Los Alamos, New Mexico 87504, USA. ³Department of Physics, Mu'tah University, Mu'tah, Karak 61710, Jordan. ⁴Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. ⁵Department of Physics and Astronomy, University of British Columbia, Vancouver V6T 1Z4, Canada. ⁶Canadian Institute for Advanced Research, Quantum Materials Program, Toronto M5G 1Z8, Canada.



in Figs 1a–c and identified from the azimuthal anisotropy in Fig. 4b), reveals the new body-centred orthorhombic Brillouin zone to be oriented concentrically with the original primitive orthorhombic Brillouin zone, and to be defined by orthogonal ordering wavevectors Q_1 and Q_2 . The

staggered twofold Fermi surface pockets at the corner T point are therefore located in the nodal region of the original Brillouin zone. A nodal Fermi surface pocket contained within the reconstructed Brillouin zone is not only consistent with the present measurements, but also with the

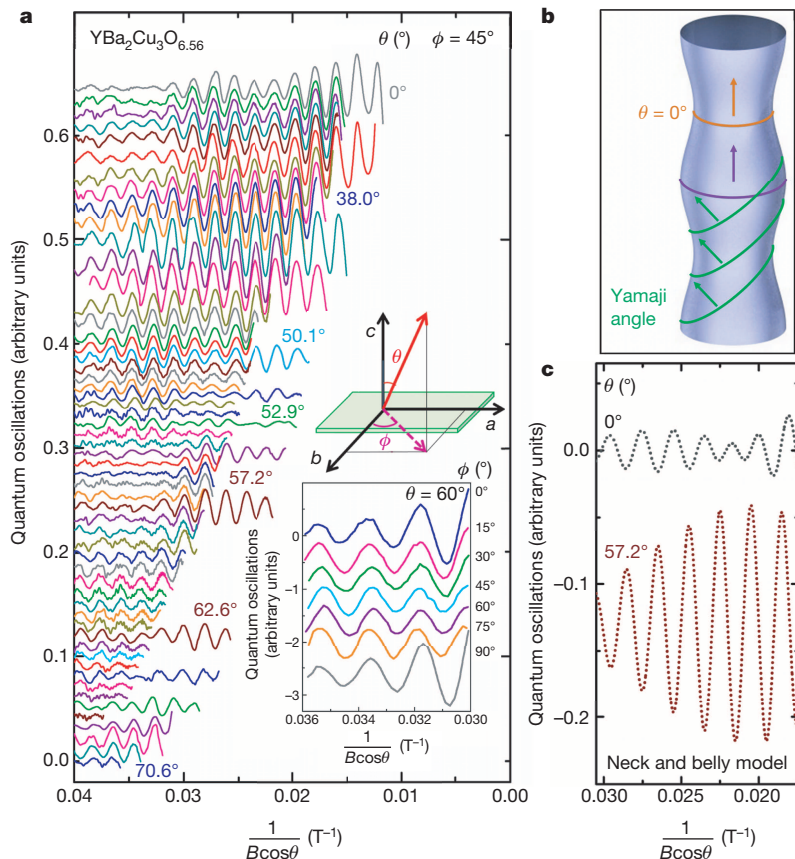
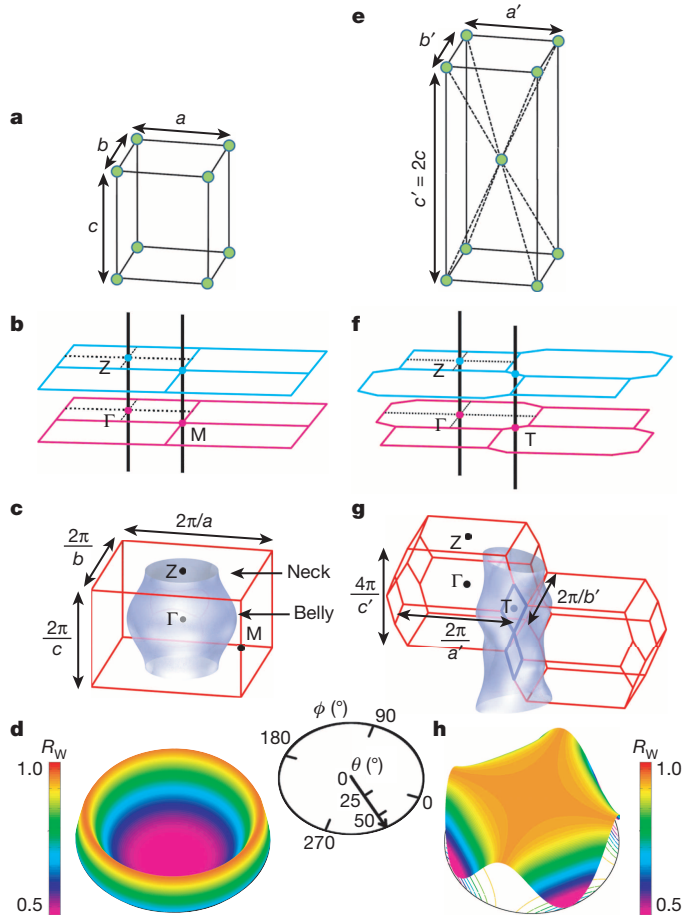


Figure 2 | Quantum oscillations defining the Fermi surface of $\text{YBa}_2\text{Cu}_3\text{O}_{6.56}$. **a**, Oscillations of the scaled contactless electrical resistivity as a function of $\frac{1}{B \cos \theta}$ for different magnetic field (B) orientations defined by ϕ and θ as indicated. The upper inset defines the angles θ and ϕ relative to the crystallographic axes. **b**, Schematic showing the degeneracy in the cyclotron orbit cross-sectional area that yields an amplitude enhancement at the Yamaji angle (green) for a neck and belly Fermi surface geometry (corrugation is accentuated throughout for clarity). **c**, Expected Yamaji resonance in the vicinity of 60° were $\text{YBa}_2\text{Cu}_3\text{O}_{6.56}$ to be described by a neck and belly Fermi surface geometry (see Methods and Extended Data Fig. 2b, c). Data in **a** and simulations in **c** have been scaled by $\exp \frac{100 \text{ T}}{B \cos \theta}$ for visual clarity.



low observed value of linear specific heat coefficient in high magnetic fields⁶ and strong chemical potential oscillations inferred from previous quantum oscillation studies¹⁷.

Figure 3 | Fermi surface and geometry-dependent quantum oscillation damping for different crystal structures. a–d, Primitive orthorhombic structure. e–h, Body-centred orthorhombic structure. Shown are the real-space unit cells (a, e), cuts through the Brillouin zone showing local symmetries about the vertical lines (b, f), location of the warped Fermi surface within the Brillouin zone in which symmetry points Γ , Z, M and T are indicated (c, g), and the angular dependences of the corresponding quantum oscillation amplitude damping factor R_w (d, h). The amplitude for the fundamental warping (d) exhibits a maximum versus θ —known as the Yamaji resonance—in contrast to an anti-resonance for the staggered twofold warping (h). R_w has been simulated using parameters in Extended Data Table 1 and a representative value of $B = 62$ T. For the definition of R_w and further analyses including the spin factor R_s , see the Methods.

We now consider the possible origin of the superlattice responsible for the above nodal staggered twofold Fermi surface pockets, which emerge from a reconstruction of a large Fermi surface determined from band structure, characteristic of the original primitive orthorhombic lattice in the normal state⁸. Numerous proposals for superlattices in the copper oxides have been put forward, for example in refs 1, 7–16 and 24–29. Of particular relevance to our observations is the charge order recently detected in short-range or long-range forms in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ by techniques such as X-ray diffraction^{19,20}, ultrasound²¹, nuclear magnetic resonance¹⁴ and optical reflectometry²². These observations point to a superlattice characterized by the same ordering wavevectors Q_1 and Q_2 identified by our measurements and defined in Fig. 1, with the superlattice scaling parameters $\delta_1 \approx \delta_2 \approx 0.3$, and anisotropic amplitudes in some cases. We note that for Fermi surface reconstruction, the superlattice need not be strictly long-range or static, but it must not be fluctuating over a range much smaller than the cyclotron radius, nor with a frequency much larger than the cyclotron frequency.

Numerical calculations indeed show that this type of charge order can give rise to nodal Fermi surface pockets similar to those observed in our measurements, potentially accompanied by antinodal gaps in the electronic excitation spectrum^{8,9,24} (schematic shown in Extended Data Figs 6 and 7). Furthermore, the resulting nodal electron-like Fermi surface pockets yield the observed negative sign of Hall coefficient in the accessed high magnetic field limit (see Methods and refs 8–10) in contrast to the positive value expected for the large unreconstructed Fermi

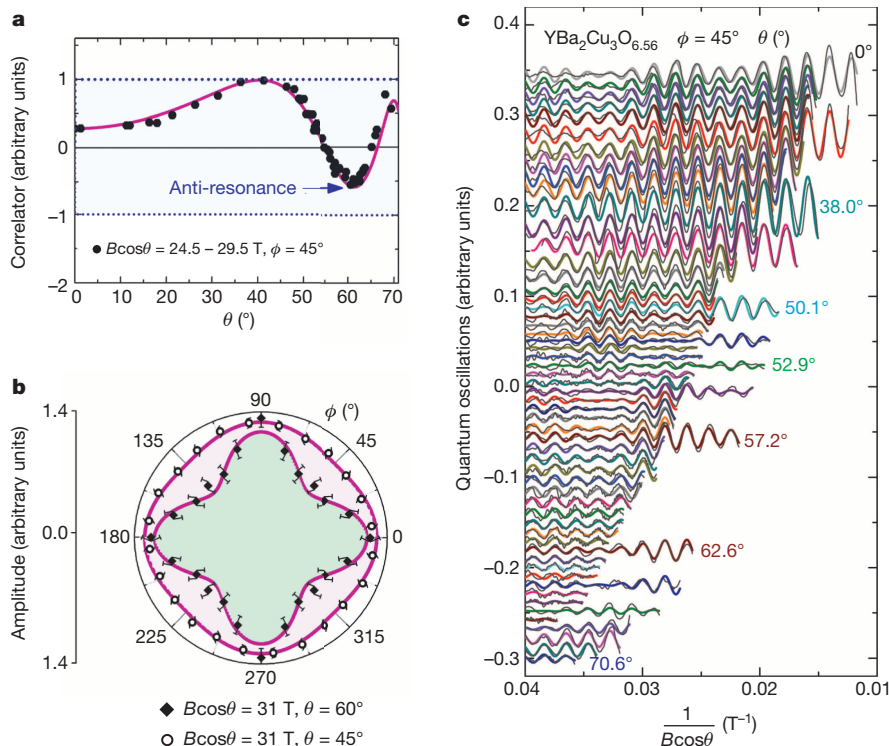


Figure 4 | Quantum oscillation data compared with a staggered twofold Fermi surface model. a, Real component of the cross-correlation between the quantum oscillation data over a fixed range of $B \cos \theta$ for a range of measured θ angles with a simple sinusoid $\cos(2\pi F/B \cos \theta + \phi)$. F and ϕ are matched to the periodicity and phase of the oscillations at $\theta = 38^\circ$, where a single frequency dominates the measured quantum oscillations. The dotted lines show the expected constant magnitude of the maximum and minimum amplitude for an ideal two-dimensional Fermi surface. A suppression in amplitude (anti-resonance) instead of a Yamaji resonance is observed as a function of θ (see Methods and Extended Data Fig. 2). b, Fourfold anisotropy in the ϕ -dependent amplitude of quantum oscillations for $\theta = 45^\circ$ and 60° . c, Measured quantum oscillations shown in Fig. 2 (black lines). The magenta lines in a and b show a simulation of R_w^{twofold} and the coloured lines in c show a simulation of Ψ^{twofold} respectively for a staggered twofold Fermi surface model (for a value of $\Delta F_{\text{twofold}} = 15$ T). For definitions of the staggered twofold damping factor R_w^{twofold} , the spin damping factor R_s , the quantum oscillation function Ψ^{twofold} , the modulation amplitude of staggered twofold geometry $\Delta F_{\text{twofold}}$ and parameter values used for the simulation, see the Methods.

surface in the absence of the superlattice. Our measurements do not distinguish between charge order with *s*-wave symmetry and unconventional symmetry such as *d*-wave symmetry (that is, bond order). Furthermore, they are compatible with charge modulation components (of wavevectors Q_1 and Q_2) not only of similar amplitudes, but also of significantly different amplitudes, as in a nearly uniaxial structure¹⁰⁷.

The strength of magnetic fields in which the present measurements were carried out (up to 85 T in Fig. 2a and up to 100 T in Extended Data Fig. 3b) is adequate to suppress long-range superconducting order, and thus to reveal the normal ground state underpinning superconductivity in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. The resulting normal ground state is a Fermi liquid characterized by staggered twofold Fermi pockets that emerge naturally from a long-range static or short-range slowly fluctuating superlattice defined by ordering wavevectors Q_1 and Q_2 (Fig. 1), support for which is found from recent complementary experiments^{14,19–21}. Superconductivity may therefore be viewed as emerging from the pairing of quasiparticles on Fermi pockets we locate in the nodal region of momentum space. Our finding clarifies observations from complementary experiments using, for example, Raman spectroscopy³⁰, which show that in the underdoped regime, Bogoliubov quasiparticles are confined to momentum space islands around the nodal regions of the Brillouin zone. This is in sharp contrast with the emergence of Bogoliubov quasiparticles from both nodal and antinodal regions of the starting large unreconstructed Fermi surface in the overdoped regime.

METHODS SUMMARY

Quantum oscillations in the electrical resistivity were measured using a contactless technique on a high quality de-twinned single crystal of $\text{YBa}_2\text{Cu}_3\text{O}_{6.56}$ over a wide range of polar and azimuthal angles, θ and ϕ , in magnetic fields up to 85 T and at 1.5 K (Fig. 2a and Extended Data Fig. 1). Measurements up to 100 T at $\theta = 0$ are presented in Extended Data Fig. 3.

The model used for the quantum oscillation simulations is discussed in the Methods. Of particular importance here is the Fermi surface geometrical damping or warping factor, taken to be of the form $R_w^{\text{rtwofold}} = J_0 \left[\frac{2\pi\Delta F_{\text{twofold}}}{B \cos \theta} \sin 2\phi J_2(k_F c' / 2) \tan \theta \right]$, which is expected for a staggered twofold Fermi surface geometry in a body-centred weakly-orthorhombic unit cell (Fig. 1a, b). Here $k_F \approx 0.13 \text{ \AA}^{-1}$ (corresponding to the dominant frequency $F_0 \approx 530 \text{ T}$) is the average radius of the Fermi surface pockets in the basal plane, $\Delta F_{\text{twofold}}$ is a measure of the magnitude of the staggered twofold warping and J_0 and J_2 are the zeroth and second-order Bessel functions, respectively. The results of the simulation are shown and explained in Fig. 4 and in Extended Data Table 1.

For a Fermi surface with fundamental neck and belly warping geometry in a primitive weakly orthorhombic unit cell (see Fig. 2b), a different geometrical damping factor $R_w^{\text{neck-belly}} = J_0 \left[\frac{2\pi\Delta F_{\text{neck-belly}}}{B \cos \theta} J_0(k_F c \tan \theta) \right]$ is required. The results of a simulation using $R_w^{\text{neck-belly}}$ with model parameters listed in Extended Data Table 3 are shown in Fig. 2c and Extended Data Fig. 2b and c. The striking difference in the angular dependence of the quantum oscillation waveform corresponding to $R_w^{\text{neck-belly}}$ and R_w^{rtwofold} (illustrated in Fig. 3d and h, respectively) is shown in Extended Data Fig. 2.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 January; accepted 2 April 2014.

Published online 15 June 2014.

- Lee, P. A., Nagaosa, N. & Wen, X. G. Doping a Mott insulator: physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78**, 17–85 (2006).
- Grissonnache, G. *et al.* Direct measurement of the upper critical field in cuprate superconductors. *Nature Commun.* **5**, 4280 (2014).
- Wang, Y. *et al.* High field phase diagram of cuprates derived from the Nernst effect. *Phys. Rev. Lett.* **88**, 257003 (2002).
- Yang, K.-Y., Rice, T. M. & Zhang, F.-C. Phenomenological theory of the pseudogap state. *Phys. Rev. B* **73**, 174501 (2006).
- Anderson, P. W. *et al.* The physics behind high-temperature superconducting cuprates: the ‘plain vanilla’ version of RVB. *J. Phys. Condens. Matter* **16**, R755–R769 (2004).
- Riggs, S. C. *et al.* Heat capacity through the magnetic-field-induced resistive transition in an underdoped high-temperature superconductor. *Nature Phys.* **7**, 332–335 (2011).

- Doiron-Leyraud, N. *et al.* Quantum oscillations and the Fermi surface in an underdoped high- T_c superconductor. *Nature* **447**, 565–568 (2007).
- Sebastian, S. E., Harrison, N. & Lonzarich, G. G. Towards resolution of the Fermi surface in underdoped high- T_c superconductors. *Rep. Prog. Phys.* **75**, 102501 (2012).
- Harrison, N. & Sebastian, S. E. Protected nodal electron pocket from multiple-Q ordering in underdoped high temperature superconductors. *Phys. Rev. Lett.* **106**, 226402 (2011).
- LeBoeuf, D. *et al.* Electron pockets in the Fermi surface of hole-doped high- T_c superconductors. *Nature* **450**, 533–536 (2007).
- Chakravarty, S. & Kee, H.-Y. Fermi pockets and quantum oscillations of the Hall coefficient in high-temperature superconductors. *Proc. Natl Acad. Sci. USA* **105**, 8835–8839 (2008).
- Millis, A. J. & Norman, M. R. Antiphase stripe order as the origin of electron pockets observed in 1/8-hole-doped cuprates. *Phys. Rev. B* **76**, 220503 (2007).
- Yao, H., Lee, D. H. & Kivelson, S. A. Fermi-surface reconstruction in a smectic phase of a high-temperature superconductor. *Phys. Rev. B* **84**, 012507 (2011).
- Wu, T. *et al.* Magnetic-field-induced charge-stripe order in the high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_y$. *Nature* **477**, 191–194 (2011).
- Chen, W.-Q., Yang, K.-Y., Rice, T. M. & Zhang, F. C. Quantum oscillations in magnetic-field-induced antiferromagnetic phase of underdoped cuprates: application to ortho-II $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$. *Europhys. Lett.* **82**, 17004 (2008).
- Sebastian, S. E. *et al.* Quantum oscillations from nodal bilayer magnetic breakdown in the underdoped high temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Phys. Rev. Lett.* **108**, 196403 (2012).
- Sebastian, S. E. *et al.* Chemical potential oscillations from nodal Fermi surface pocket in the underdoped high-temperature superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. *Nature Commun.* **2**, 471 (2011).
- Hossain, M. A. *et al.* In situ doping control of the surface of high-temperature superconductors. *Nature Phys.* **4**, 527–531 (2008).
- Ghiringhelli, G. *et al.* Long-range incommensurate charge fluctuations in $(\text{Y,Nd})\text{Ba}_2\text{Cu}_3\text{O}_{6+x}$. *Science* **337**, 821–825 (2012).
- Chang, J. *et al.* Direct observation of competition between superconductivity and charge density wave order in $\text{YBa}_2\text{Cu}_3\text{O}_{6.67}$. *Nature Phys.* **8**, 871–876 (2012).
- LeBoeuf, D. *et al.* Thermodynamic phase diagram of static charge order in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_y$. *Nature Phys.* **9**, 79–83 (2013).
- Hinton, J. P. *et al.* A new collective mode in YBCO observed by time-domain reflectometry. *Phys. Rev. B* **88**, 060508 (2013).
- Bergemann, C., Mackenzie, A. P., Julian, S. R., Forsythe, D. & Ohmichi, E. Quasi-two-dimensional Fermi liquid properties of the unconventional superconductor Sr_2RuO_4 . *Adv. Phys.* **52**, 639–725 (2003).
- Li, J.-X., Wu, C.-Q. & Lee, D.-H. Checkerboard charge density wave and pseudogap of high- T_c cuprate. *Phys. Rev. B* **74**, 184515 (2006).
- Wang, Y. & Chubukov, A. V. Charge order and loop currents in hole-doped cuprates. Preprint at <http://arxiv.org/abs/1401.0712> (2014).
- Castellani, C., Di Castro, C. & Grilli, M. Singular quasiparticle scattering in the proximity of charge instabilities. *Phys. Rev. Lett.* **75**, 4650–4653 (1995).
- Hayward, L. E., Hawthorn, D. G., Melko, R. G. & Sachdev, S. Angular fluctuations of a multi-component order describe the pseudogap regime of the cuprate superconductors. *Science* **343**, 1336–1339 (2014).
- Eftov, K. B., Meier, H. & Pépin, C. Pseudogap state near a quantum critical point. *Nature Phys.* **9**, 442–446 (2013).
- Nie, L., Tarjus, G. & Kivelson, S. A. Quenched disorder and vestigial nematicity in the pseudo-gap regime of the cuprates. *Proc. Natl Acad. Sci.* **111**, 7980–7985 (2014).
- Sacuto, A. *et al.* New insights into the phase diagram of the copper oxide superconductors from electronic Raman scattering. *Rep. Prog. Phys.* **76**, 022502 (2013).

Acknowledgements S.E.S. acknowledges support from the Royal Society, King’s College Cambridge, the Winton Programme for the Physics of Sustainability, and the European Research Council under the European Union’s Seventh Framework Programme (grant number FP/2007–2013)/ERC Grant Agreement number 337425-SUPERCONDUCTINGMOTT. N.H. and F.F.B. acknowledge support for high-magnetic-field experiments from the US Department of Energy, Office of Science, BES-MSE ‘Science of 100 Tesla’ programme. G.G.L. acknowledges support from Engineering and Physical Sciences Research Council (EPSRC) grant EP/K012894/1. P.A.G. is supported by the EPSRC and thanks the University of Oxford for the provision of a Visiting Lectureship. R.L., D.A.B. and W.N.H. acknowledge support from the Canadian Institute for Advanced Research, and the Natural Science and Engineering Research Council. A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by NSF co-operative agreement number DMR-0654118, the state of Florida, and the DOE. We acknowledge discussions with many colleagues, including H. Alloul, C. Bergemann, A. Carrington, S. Chakravarty, A. Chubukov, E. M. Forgan, S. R. Julian, B. Keimer, S. A. Kivelson, R. B. Laughlin, M. Le Tacon, L. Taillefer, D.-H. Lee, P. A. Lee, P. B. Littlewood, A. P. Mackenzie, M. R. Norman, C. Pépin, C. Proust, M. Randeria, S. Sachdev, A. Sacuto, T. Senthil, J. P. Sethna, J. Tranquada and C. M. Varma. We are grateful for the experimental support provided by the ‘100 T’ team, including J. B. Betts, Y. Coulter, M. Gordon, C. H. Mielke, A. Parish, D. Rickel and D. Roybal.

Author Contributions S.E.S., N.H., F.F.B., M.M.A. and P.A.G. performed high magnetic field measurements. R.L., D.A.B. and W.N.H. prepared single crystals. S.E.S., N.H. and G.G.L. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.S. (suchitra@phy.cam.ac.uk) or N.H. (nharrison@lanl.gov).

Giant nonlinear response from plasmonic metasurfaces coupled to intersubband transitions

Jongwon Lee¹, Mykhailo Tymchenko¹, Christos Argyropoulos¹, Pai-Yen Chen¹, Feng Lu¹, Frederic Demmerle², Gerhard Boehm², Markus-Christian Amann², Andrea Alù¹ & Mikhail A. Belkin¹

Intersubband transitions in n-doped multi-quantum-well semiconductor heterostructures make it possible to engineer one of the largest known nonlinear optical responses in condensed matter systems—but this nonlinear response is limited to light with electric field polarized normal to the semiconductor layers^{1–7}. In a different context, plasmonic metasurfaces (thin conductor–dielectric composite materials) have been proposed as a way of strongly enhancing light–matter interaction and realizing ultrathin planarized devices with exotic wave properties^{8–11}. Here we propose and experimentally realize metasurfaces with a record-high nonlinear response based on the coupling of electromagnetic modes in plasmonic metasurfaces with quantum-engineered electronic intersubband transitions in semiconductor heterostructures. We show that it is possible to engineer almost any element of the nonlinear susceptibility tensor of these structures, and we experimentally verify this concept by realizing a 400-nm-thick metasurface with nonlinear susceptibility of greater than 5×10^4 picometres per volt for second harmonic generation at a wavelength of about 8 micrometres under normal incidence. This susceptibility is many orders of magnitude larger than any second-order nonlinear response in optical metasurfaces measured so far^{12–15}. The proposed structures can act as ultrathin highly nonlinear optical elements that enable efficient frequency mixing with relaxed phase-matching conditions, ideal for realizing broadband frequency up- and down-conversions, phase conjugation and all-optical control and tunability over a surface.

Many demonstrated applications of optical metamaterials are based on their linear interaction with light, including super-resolution imaging^{16,17} and optical cloaking^{18–20}. More recently, optical metamaterials with tailored nonlinear response have provided new degrees of freedom in metamaterial design, with interesting uses in super-resolution imaging²¹, performing efficient frequency conversion and optical control with greatly relaxed phase-matching conditions²², and in optical switching and memories at the nanoscale²³.

So far, nonlinearities in metamaterials have been mostly realized by exploiting the natural nonlinear response of plasmonic metals^{12,13} or by enhancing the nonlinearity of optical crystals using plasmonic nanoantennas^{14,15}. A different approach to realizing large nonlinear optical response has been put forward—quantum-engineering electronic intersubband transitions in n-doped multi-quantum-well (MQW) semiconductor heterostructures^{1–7}. By controlling the widths of wells and barriers in the MQW structures, one can tailor the transition energy and dipole moments between electron subbands so as to maximize the quantum-mechanical expression for a nonlinear process of choice; this process can produce one of the largest known nonlinear responses, up to six orders of magnitude larger than that of traditional nonlinear optical materials^{1–7}. Voltage may be used to modify and spectrally tune intersubband nonlinearities³, and electrical pumping may be used to produce active intersubband structures with full loss-compensation for both second-order^{5–7} and third-order²⁴ nonlinear processes.

Nonlinear MQW structures have been successfully integrated into waveguide-based systems so as to produce efficient frequency conversion^{2,4–7,25},

and have enabled the development of mass-producible room-temperature electrically pumped sources of THz radiation⁷. A power conversion efficiency of nearly 1% for second harmonic generation (SHG) at a 8.6 μm fundamental frequency was achieved in waveguides with passive $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ MQW structures²⁵ and a power conversion efficiency of over 16% was theoretically predicted²⁶. However, the integration of giant MQW nonlinearities with free-space optics is very challenging, because optical transitions between electron subbands are intrinsically polarized along the surface normal to the MQW layers (taken to be the z axis in this Letter).

In a different context, properly patterned metallic surfaces supporting highly confined plasmonic resonances have been proposed for a variety of applications, including enhanced light–matter interaction at the subwavelength scale, polarization conversion, control of optical transmission and radiation, enhanced chirality, asymmetric transmission and filtering¹¹. In order to overcome the present limitations of MQW systems, and further enhance the intersubband nonlinear response, we propose combining the MQW layered substrates with suitably designed plasmonic metasurfaces. In the following we theoretically and experimentally demonstrate that, by combining quantum-electronic engineering of intersubband nonlinearities with electromagnetic engineering of plasmonic nanoresonators, we can produce ultrathin, planarized, highly nonlinear optical metasurfaces. This approach allows us to create large-area metasurfaces in which virtually any element of the nonlinear susceptibility tensor may be *ad hoc* engineered to have a giant nonlinear response. The impact of these ultrathin devices may be substantial in a variety of fields, including THz generation and detection, phase conjugation, and other nonlinear processes.

Figure 1a shows the band diagram of a coupled quantum well structure, repeated multiple times to produce the 400-nm-thick MQW layer used in our experiments. This structure is designed to support a giant resonant nonlinear response for SHG at $1/\lambda \approx 1,240 \text{ cm}^{-1}$ ($\lambda \approx 8 \mu\text{m}$). Figure 1b plots the absolute value of the SHG nonlinear susceptibility tensor element $\chi_{zzz}^{(2)}$ as a function of pump frequency calculated for this structure as described in Methods. The nonlinear response peaks at approximately 54 nm V^{-1} , nearly 3 orders of magnitude larger than the largest $\chi^{(2)}$ coefficient of natural optical materials²⁷.

The MQW layer is then sandwiched between a metal ground plane and a patterned array of metallic nanostructures (Fig. 1c, d). This grounded metasurface can be designed to enable efficient coupling of plasmonic modes with z -polarized intersubband transitions, drastically enhancing light–matter interaction and converting the impinging transverse field polarization to the desired z -direction²⁸. Even though a ground plane is used in our metasurface design, efficient coupling of intersubband transitions with resonant metallic nanostructures may also be achieved without using a ground plane, as shown recently²⁹, which may allow the results presented here to be translated into transmission mode.

Since our goal is to enhance SHG, the plasmonic nanostructures are designed to induce and enhance the local electric field at both fundamental frequency (FF) ω and the second-harmonic frequency (SH) 2ω . The plasmonic nanostructures must also lack inversion symmetry

¹Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712, USA. ²Walter Schottky Institut, Technische Universität München, Am Coulombwall 4, Garching 85748, Germany.

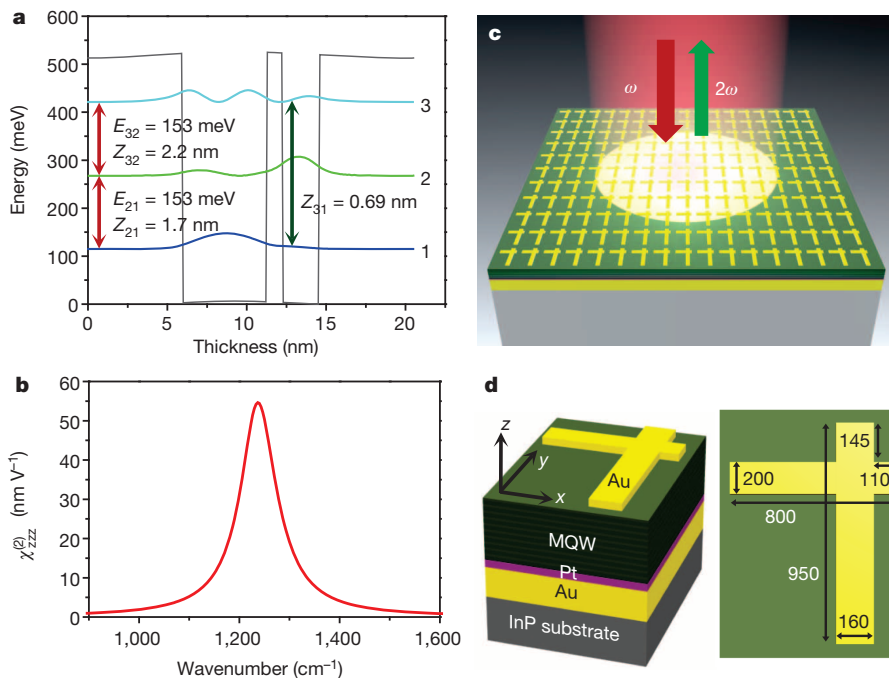


Figure 1 | Nonlinear metasurface structure. **a**, Conduction band diagram of one period of an $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ coupled quantum well structure designed for giant nonlinear response for SHG. The moduli squared of the electron wavefunctions for subbands 1, 2 and 3 are shown and labelled accordingly. Transitions between pairs of electron subbands are marked with double-headed red arrows, and the values of the transition energies (E_{21} and E_{32}) and dipole moments (Z_{21} , Z_{32} and Z_{31}) are shown next to each arrow. **b**, Intersubband nonlinear susceptibility of the structure in **a** as a function of pump frequency for SHG. **c**, Schematic of the metasurface design and operation. Red and green arrows indicate the incident pump beam at fundamental frequency ω and the reflected second-harmonic beam at frequency 2ω , respectively. **d**, A $1,000 \text{ nm} \times 1,300 \text{ nm}$ metasurface unit cell (left), dimensions of the nanocross are given in nm (right).

in the x - y plane, otherwise all nonlinear susceptibility tensor elements $\chi_{ijk}^{(2)}$ with i, j and k being x or y would vanish by symmetry.

Our starting design strategy consists of an array of 'L'-shaped nanostructures, whose two arms have been shown in recent papers to support two independently tunable plasmonic resonances that can be used for polarization control^{30,31}. After several rounds of optimization, we converged on the nanostructure shown in Fig. 1d, consisting of an asymmetric nanocross with the two orthogonal arms connected in such a way as to ensure good modal overlap between the resonances at ω and 2ω , yet possessing a strong asymmetry to support large $\chi_{ijk}^{(2)}$. The dimensions of the nanocross optimized for SHG at $1/\lambda = 1,240 \text{ cm}^{-1}$ are shown in Fig. 1d and the results of our full-wave electromagnetic simulations are shown in Fig. 2. In particular, Fig. 2a-h shows the normalized E_z field component induced in the MQW layer for different FF and SH input light polarizations at normal incidence. The computed fields are normalized to the input (transverse) field amplitude impinging at normal incidence. Significant E_z field is induced in the MQW at both ω and 2ω , up to five times the amplitude of the transverse impinging field.

The effective second-order nonlinear susceptibility tensor of the metasurface ($\tilde{\chi}^{(2)\text{eff}}$) may be related to these field distributions by using the Lorentz reciprocity theorem³², as described in the Methods section, yielding the expression for the general ijk element

$$\chi_{ijk}^{(2)\text{eff}} = \chi_{zzz}^{(2)} \frac{\int_V E_{z(i)}^{2\omega}(x, y, z) E_{z(j)}^{\omega}(x, y, z) E_{z(k)}^{\omega}(x, y, z) dV}{E_{i(\text{inc})}^{2\omega} E_{j(\text{inc})}^{\omega} E_{k(\text{inc})}^{\omega} V} \quad (1)$$

where $E_{z(i)}^{\omega}$ and $E_{z(i)}^{2\omega}$ are the local E_z field components in the MQW layer induced by i -polarized incident waves $E_{i(\text{inc})}^{\omega}$ and $E_{i(\text{inc})}^{2\omega}$, respectively, at frequencies ω and 2ω , respectively, i can be x, y or z , and the integration covers the volume (V) of the MQW layer in the unit cell.

Equation (1) shows that we can engineer essentially any component of the nonlinear susceptibility tensor of the metasurface and further increase the already giant nonlinear response from intersubband transitions using local field enhancement at ω and 2ω . It also confirms the necessity of inducing a strong modal overlap between the fundamental and SH modes in the nanostructures, for which our metasurface is optimized. This result may be readily generalized to other nonlinear processes of interest. Using equation (1) and simulated field distributions for the nanocross structure in Fig. 2, we calculate that the largest nonlinear susceptibility coefficient for normal light incidence at an input wavenumber

of $1,240 \text{ cm}^{-1}$ is $\chi_{yyy}^{(2)\text{eff}} = 31 \text{ nm V}^{-1}$, followed by $\chi_{xxx}^{(2)\text{eff}} = 25 \text{ nm V}^{-1}$, $\chi_{xyy}^{(2)\text{eff}} = 6.5 \text{ nm V}^{-1}$ and $\chi_{yxx}^{(2)\text{eff}} = 3.9 \text{ nm V}^{-1}$.

Figure 2i shows the computed linear absorption spectrum of the metasurface. The calculated SH power versus FF pump power for different input/output polarization combinations is shown in Fig. 2j. Results are given for yyy , xxx , xyy and yxx polarization combinations, where the first index refers to the polarization of the SH beam, and the last two indices refer to the polarization of the FF beam.

For experimental testing, a 400-nm-thick MQW layer was transferred onto a metal ground plane and a $400 \mu\text{m} \times 400 \mu\text{m}$ two-dimensional array of nanocrosses was fabricated on top of the MQW layer to produce the structure shown in Fig. 1c. Fabrication details are provided in Methods. To accommodate the uncertainty in material parameters and fabrication errors, we fabricated several arrays with nanocross dimensions slightly varying around the optimal theoretical parameters shown in Fig. 1d. The description below focuses on the measured array with highest SHG generation efficiency at $1,240 \text{ cm}^{-1}$, but we stress that the measured nonlinear response is quite robust to small changes in the design parameters, as several of our realized metasurface samples showed a similarly large nonlinear response. Yet we verified that metasurfaces with plasmonic resonances significantly detuned from the intersubband transition frequencies of the MQW structure produced, as expected, little or no nonlinear response (see Methods for details). This confirms the unique property of the proposed metasurface, combining quantum and electromagnetic engineering of its nonlinear response.

Figure 3 shows a scanning electron microscope (SEM) image of the nanocross array (Fig. 3a) and a side view of the cleaved facet with the MQW layer metal-bonded to the InP substrate (Fig. 3b). The absorption spectra of the metasurface for x - and y -polarized input light at normal incidence are shown in Fig. 3c. The experimental shapes and positions of the peaks are in reasonable agreement with the simulated spectrum in Fig. 2i.

SHG measurements of the metasurface were performed using the set-up shown in Fig. 3d and described in the figure legend. SHG peak power as a function of the square of the FF peak power and the square of the FF peak intensity is plotted in Fig. 4a for different input/output polarization combinations. The highest SHG generation efficiency is achieved for the yyy polarization combination, followed by xxx , yxx and xyy polarization combinations, in agreement with the theoretical predictions shown in Fig. 2j. Comparably strong nonlinear responses were also observed for the same metasurface for a 45° incidence angle geometry, indicating that

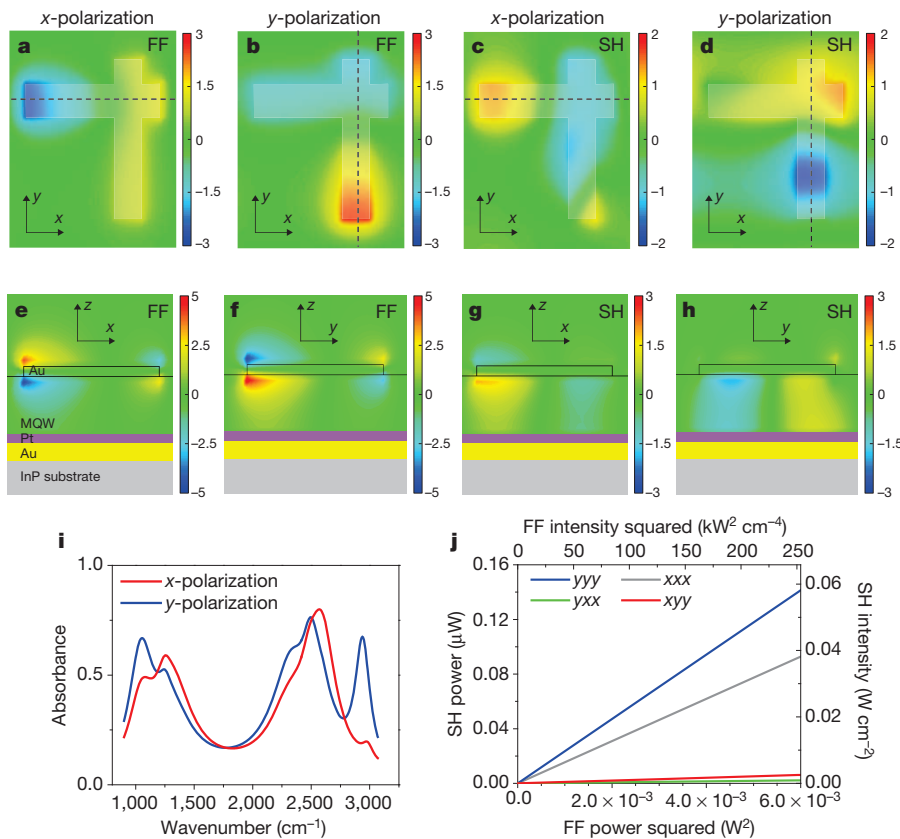


Figure 2 | Simulations and predicted performance of the metasurface geometry. **a–d**, Top view cross-section of the calculated E_z field enhancement monitored in the MQW layer 100 nm below the gold plasmonic resonators at FF (**a**, **b**) and SH frequency (**c**, **d**). **e–h**, Side view cross-section of the calculated E_z field enhancement monitored along the dashed lines shown in **a–d** respectively at FF (**e**, **f**) and SH frequency (**g**, **h**). The data are shown for incident light polarized along the x -axis and the y -axis of the structure. **i**, Computed absorption spectrum of the proposed metasurface for different input light polarization. **j**, SHG power (left axis) and intensity (right axis) versus the square of the FF power input/output polarization combinations for different input/output polarization combinations for FF wavenumber $1/\lambda = 1,240 \text{ cm}^{-1}$. The FF focal spot is assumed to be of a Gaussian shape with a diameter of $2w = 35 \mu\text{m}$, as in the experiments (Methods).

our structures may be operated without constraints on the incidence and observation angles (see Methods for details).

From equation (1) and the detailed analysis in the Methods section, one expects to observe a linear dependence of SH power on the square of FF power in Fig. 4a. However, as indicated in the figure, the slope of the curve changes: for example, for the yyy polarization combination, the slope changes from $57 \mu\text{W W}^{-2}$ for low FF powers to $23 \mu\text{W W}^{-2}$ for higher FF powers. We attribute this effect to intensity saturation of intersubband transitions³³ in our MQW structures, and provide further experimental evidence in support of this conclusion in Methods. Calculations shown in Methods indicate that the saturation intensity³³ for the state 1 to state 2 transition (see Fig. 1) in our structures is approximately 0.47 MW cm^{-2} , and simulation results in Fig. 2 indeed indicate local intensities approaching saturation intensity in the hotspots of the MQW structure for input pump intensities above 10 kW cm^{-2} . We note that our proof-of-concept

metasurface design is far from perfect, as one ideally wants to create a more uniform local field enhancement within the MQW structure at both FF and SH frequencies with maximal modal overlap integrals, consistent with equation (1). This implies that an even larger nonlinear response may be found with better optimized nanoresonator designs, exploiting the field localization properties of plasmonic structures. Yet the measured results presented here reveal an enormous potential for the proposed metasurface.

To better understand the potential improvement associated with nanoresonator optimization, consider an ideal metasurface in which the nanoresonators are able to induce a uniform E_z field enhanced 3 times compared to the impinging field across the entire MQW layer at both SH and FF. In this case, equation (1) predicts $\chi^{(2)\text{eff}}_{zzz} = 27\chi^{(2)}_{zzz}$, approximately 50 times larger than the largest $\chi^{(2)\text{eff}}$ coefficients computed for the nanocross design used in this work. Since the SHG power scales with $|\chi^{(2)\text{eff}}|^2$, such a metasurface would produce over three orders of magnitude

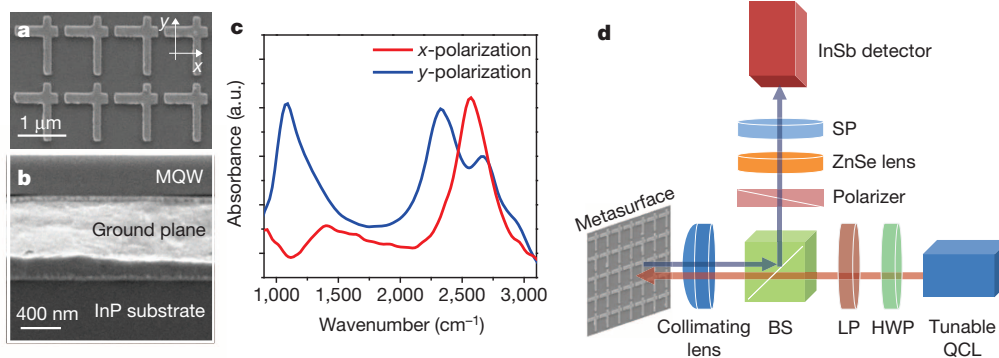


Figure 3 | Characterization of processed metasurface. **a**, **b**, Scanning electron microscope images of the fabricated metasurface, top (**a**) and side (**b**) view. **c**, Absorption spectrum of the fabricated metasurface for normally incident light polarized along x -axis and y -axis of nanocrosses, as shown in **a**. **d**, Optical set-up for metasurface characterization. Linearly polarized light from a tunable quantum cascade laser passes through an achromatic half-wave

plate (HWP) for polarization control, a long pass filter (LP) to remove SH light coming from the laser, a non-polarizing achromatic 50/50 beam splitter (BS), and a numerical aperture 0.5 collimating lens to the sample. SHG output is collected by the same lens and is directed by the beam splitter towards the detector through a polarizer, a ZnSe lens, and a short pass filter (SP) that is used to block FF.

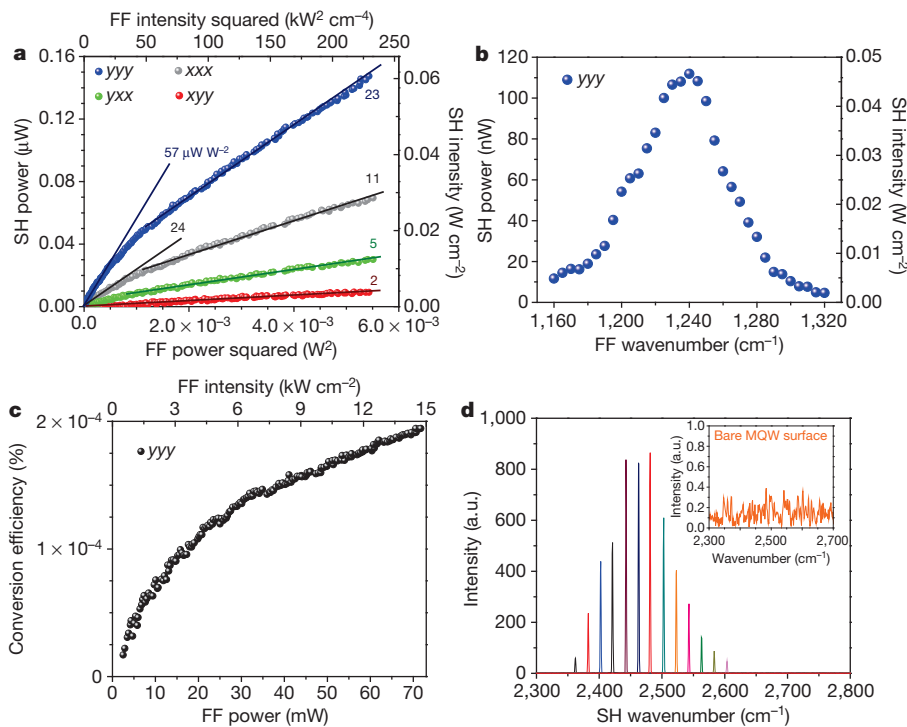


Figure 4 | Nonlinear response from the metasurface. **a**, SH peak power (left axis) and intensity (right axis) output as a function of FF peak power squared (bottom axis) or peak intensity squared (top axis) at an FF wavenumber of $1,240\text{ cm}^{-1}$ for different input/output polarization combinations (see key). Straight lines and labels indicate slopes of SH peak power dependence on FF power squared. **b**, SH peak power (left axis) and peak intensity (right axis) as a function of FF wavenumber for the yyy polarization combination. FF peak power was fixed at 30 mW for the measurements. **c**, SHG power conversion efficiency as a function of FF peak power (bottom axis) or peak intensity (top axis) at an FF wavenumber of $1,240\text{ cm}^{-1}$ for the yyy polarization combination. **d**, Spectra of SH output for different pump wavenumbers. Inset, no SH emission is observable for a bare MQW surface with no nanostructures on it. Measurements in **a–d** are performed at normal FF incidence/normal SH reflection.

improvement in SHG conversion efficiency compared to our current experiment. Furthermore, since the field enhancement is limited to a factor of 3, no significant intensity saturation is expected for the pump intensities used in our experiment. Improvements may also be achieved by exploring the quantum-engineering design space of our structures. The simple doubly-resonant quantum well heterostructure used here for this initial proof-of-concept demonstration is by no means the optimal design, and more sophisticated designs may be devised to achieve both higher nonlinearity and much larger saturation intensity, similar to those described in refs 25 and 26.

Our MQW structure and plasmonic nanocrosses were designed to provide maximum SHG efficiency at a pump wavenumber of $1/\lambda \approx 1,240\text{ cm}^{-1}$. This is experimentally confirmed in Fig. 4b, where we record SHG power output as a function of pump frequency while keeping the pump power constant. The yyy SHG power conversion efficiency for an FF wavenumber of $1,240\text{ cm}^{-1}$ is shown in Fig. 4c. We achieve a power conversion efficiency of almost 2×10^{-6} using a pump intensity of only 15 kW cm^{-2} . In comparison, previously demonstrated nonlinear optical metasurfaces at infrared or visible frequencies required high-peak-power fs lasers providing over 10^8 W cm^{-2} FF intensity to produce an SHG power conversion efficiency of 10^{-10} to 10^{-11} (refs 12–15). Assuming identical low pumping intensity, our metasurfaces achieve over 8 orders of magnitude enhancement in conversion efficiency over the current state-of-the-art.

Experimental values of $\chi_{ijk}^{(2)\text{eff}}$ are obtained using the intensity data for SH and FF beams in Fig. 4 as described in Methods. We obtain $\chi_{yyy}^{(2)\text{eff}} \approx 55\text{ nm V}^{-1}$ (35 nm V^{-1}) and $\chi_{xxx}^{(2)\text{eff}} \approx 36\text{ nm V}^{-1}$ (24 nm V^{-1}) for low FF intensity (high FF intensity) in Fig. 4a. The other two coefficients have no significant dependence on pump intensity in the experiment: $\chi_{yxx}^{(2)\text{eff}} \approx 16\text{ nm V}^{-1}$ and $\chi_{xyy}^{(2)\text{eff}} = 10\text{ nm V}^{-1}$. All these quantities are in good agreement with our theoretical predictions, given uncertainties in material parameters and $\chi_{zzz}^{(2)}$. Finally, Fig. 4d shows SHG emission spectra from the metasurface for different FF. The MQW structure without patterned nanostructures does not produce any significant nonlinear response for normal incidence, as shown in the figure inset.

Our proof-of-concept experiment may be extended in a variety of ways: for example, to create voltage-tunable³ and active electrically pumped nonlinear metasurfaces^{5–7}, and to spectrally and spatially engineer the nonlinear optical response in more general ways. The metasurfaces with

giant nonlinear optical response presented here may serve as the foundation for a flat nonlinear optics paradigm, in which efficient frequency mixing may occur over deeply subwavelength films with significantly relaxed phase-matching conditions (compared to those required for bulk nonlinear crystals) that only require matching of wavevector components parallel to the surface. Such ultrathin nonlinear materials may dramatically simplify wave-mixing experiments in a variety of set-ups, and find application in frequency up- and down-conversion, mixing, phase conjugation, all-optical control and tunability, as well as in photon pair generation over a surface for quantum information processing.

METHODS SUMMARY

The $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ coupled quantum well structure for nonlinear response shown in Fig. 1a was designed using a self-consistent Poisson-Schrodinger solver. The layer sequence (in nanometres) is **6.0/5.3/1/2.3/6.0**, where AlInAs barriers are shown in bold, and the first 3 nm of the first 6-nm barrier and the last 3 nm of the last 6-nm barrier are n-doped to $5 \times 10^{17}\text{ cm}^{-3}$. A 577-nm-thick MQW layer composed of 28 repetitions of the structure in Fig. 1a was grown by molecular beam epitaxy on a semi-insulating InP substrate. The layer was then transferred to another gold-coated semi-insulating InP substrate via thermo-compression bonding and thinned down to approximately 400 nm thickness by chemical etching. A commercial Maxwell's equations solver based on the finite-integration method was used for the electromagnetic simulations. Experimentally, a $400\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$ two-dimensional array of plasmonic nanoresonators was patterned onto the top of the MQW layer via e-beam lithography, metal evaporation, and lift-off. Linear optical characterization of the metasurface was done in reflection mode using a Fourier-transform infrared spectrometer (FTIR) equipped with a liquid-nitrogen-cooled HgCdTe photodetector. Nonlinear optical characterization was done using the optical set-up shown in Fig. 3d. A pulsed broadly-tunable quantum cascade laser was used for pumping and a calibrated InSb detector was used for SHG signal detection. We assume a Gaussian beam shape on the metasurface with an intensity distribution for FF given as $I_{\text{FF}}e^{-2r^2/w^2}$ and that for SH given as $I_{\text{SH}}e^{-4r^2/w^2}$. The FF focal spot diameter was measured by the knife-edge technique to be $2w = 35\text{ }\mu\text{m}$.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2013; accepted 9 May 2014.

- Fejer, M. M. *et al.* Observation of extremely large quadratic susceptibility at $9.6\text{--}10.8\text{ }\mu\text{m}$ in electric-field-biased AlGaAs quantum wells. *Phys. Rev. Lett.* **62**, 1041–1044 (1989).

2. Rosencher, E., Bois, P., Nagle, J. & Delaire, S. Second harmonic generation by intersubband transitions in compositionally asymmetrical MQWs. *Electron. Lett.* **25**, 1063–1065 (1989).
3. Capasso, F., Sirtori, C. & Cho, A. Y. Coupled-quantum-well semiconductors with giant electric-field tunable nonlinear-optical properties in the infrared. *IEEE J. Quantum Electron.* **30**, 1313–1326 (1994).
4. Rosencher, E. *et al.* Quantum engineering of optical nonlinearities. *Science* **271**, 168–173 (1996).
5. Gmachl, C. *et al.* Optimized second-harmonic generation in quantum cascade lasers. *IEEE J. Quantum Electron.* **39**, 1345–1355 (2003).
6. Belkin, M. A. *et al.* Terahertz quantum-cascade-laser source based on intracavity difference-frequency generation. *Nature Photon.* **1**, 288–292 (2007).
7. Vijayraghavan, K. *et al.* Broadly tunable terahertz generation in mid-infrared quantum cascade lasers. *Nature Commun.* **4**, 2021 (2013).
8. Liu, X. L. *et al.* Taming the blackbody with infrared metamaterials as selective thermal emitters. *Phys. Rev. Lett.* **107**, 045901 (2011).
9. Yu, N. F. *et al.* Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334**, 333–337 (2011).
10. Ni, X. J. *et al.* Broadband light bending with plasmonic nanoantennas. *Science* **335**, 427 (2012).
11. Monticone, F., Estakhri, N. M. & Alu, A. Full control of nanoscale optical transmission with a composite metascreen. *Phys. Rev. Lett.* **110**, 203903 (2013).
12. Klein, M. W., Enkrich, C., Wegener, M. & Linden, S. Second-harmonic generation from magnetic metamaterials. *Science* **313**, 502–504 (2006).
13. Feth, N. *et al.* Second-harmonic generation from complementary split-ring resonators. *Opt. Lett.* **33**, 1975–1977 (2008).
14. Fan, W. J. *et al.* Second harmonic generation from patterned GaAs inside a subwavelength metallic hole array. *Opt. Express* **14**, 9570–9575 (2006).
15. Niesler, F. B. P. *et al.* Second-harmonic generation from split-ring resonators on a GaAs substrate. *Opt. Lett.* **34**, 1997–1999 (2009).
16. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
17. Fang, N., Lee, H., Sun, C. & Zhang, X. Sub-diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534–537 (2005).
18. Leonhardt, U. Optical conformal mapping. *Science* **312**, 1777–1780 (2006).
19. Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
20. Chen, P. Y., Soric, J. & Alu, A. Invisibility and cloaking based on scattering cancellation. *Adv. Mater.* **24**, Op281–Op304 (2012).
21. Pendry, J. B. Time reversal and negative refraction. *Science* **322**, 71–73 (2008).
22. Rose, A. & Smith, D. R. Overcoming phase mismatch in nonlinear metamaterials. *Opt. Mater. Express* **1**, 1232–1243 (2011).
23. Argyropoulos, C. *et al.* Nonlinear plasmonic cloaks to realize giant all-optical scattering switching. *Phys. Rev. Lett.* **108**, 263905 (2012).
24. Hugi, A. *et al.* Mid-infrared frequency comb based on a quantum cascade laser. *Nature* **492**, 229–233 (2012).
25. Vodopyanov, K. L. *et al.* Phase-matched second harmonic generation in asymmetric double quantum wells. *Appl. Phys. Lett.* **72**, 2654–2656 (1998).
26. Vurgaftman, I., Meyer, J. R. & RamMohan, L. R. Optimized second-harmonic generation in asymmetric double quantum wells. *IEEE J. Quantum Electron.* **32**, 1334–1346 (1996).
27. Boyd, R. W. *Nonlinear Optics* (Academic, 2008).
28. Todorov, Y. *et al.* Ultrastrong light-matter coupling regime with polariton dots. *Phys. Rev. Lett.* **105**, 196402 (2010).
29. Benz, A. *et al.* Strong coupling in the sub-wavelength limit using metamaterial nanocavities. *Nature Commun.* **4**, 2882 (2013).
30. Ellenbogen, T., Seo, K. & Crozier, K. B. Chromatic plasmonic polarizers for active visible color filtering and polarimetry. *Nano Lett.* **12**, 1026–1031 (2012).
31. Zhao, Y. & Alu, A. Tailoring the dispersion of plasmonic nanorods to realize broadband optical meta-waveplates. *Nano Lett.* **13**, 1086–1091 (2013).
32. Balanis, C. A. *Advanced Engineering Electromagnetics* (Wiley, 1989).
33. Vodopyanov, K. L. *et al.* Intersubband absorption saturation study of narrow III–V multiple quantum wells in the $\lambda = 2.8\text{--}9\text{ }\mu\text{m}$ spectral range. *Semicond. Sci. Technol.* **12**, 708–714 (1997).

Acknowledgements This work was supported by NSF EAGER grant no. 1348049 (to M.A.B. and A.A.), AFOSR YIP award no. FA9550-10-1-0076 (M.A.B.), AFOSR YIP award no. FA9550-11-1-0009 (A.A.), and ONR MURI grant no. N00014-10-1-0942 (A.A.). The Walter Schottky Institute group acknowledges support from the Excellence Cluster 'Nano Initiative Munich (NIM)'. Sample fabrication was carried out in the Microelectronics Research Center at the University of Texas at Austin, which is a member of the National Nanotechnology Infrastructure Network.

Author Contributions J.L. designed the semiconductor heterostructure, calculated physical parameters and performed all fabrication and experimental measurements; M.T., C.A. and P.-Y.C. performed theoretical computations and structure optimization; F.L. assisted in experimental measurements; F.D., G.B. and M.-C.A. performed the semiconductor heterostructure growth; M.A.B. conceived the concept and the experiment; M.A.B. and A.A. developed the concept and planned and directed the research; and J.L., M.T., C.A., A.A. and M.A.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.B. (mbelkin@ece.utexas.edu).

Quantum control and process tomography of a semiconductor quantum dot hybrid qubit

Dohun Kim¹, Zhan Shi¹, C. B. Simmons¹, D. R. Ward¹, J. R. Prance¹, Teck Seng Koh¹, John King Gamble², D. E. Savage³, M. G. Lagally³, Mark Friesen¹, S. N. Coppersmith¹ & Mark A. Eriksson¹

The similarities between gated quantum dots and the transistors in modern microelectronics^{1,2}—in fabrication methods, physical structure and voltage scales for manipulation—have led to great interest in the development of quantum bits (qubits) in semiconductor quantum dots^{3–18}. Although quantum dot spin qubits have demonstrated long coherence times, their manipulation is often slower than desired for important future applications, such as factoring¹⁹. Furthermore, scalability and manufacturability are enhanced when qubits are as simple as possible. Previous work has increased the speed of spin qubit rotations by making use of integrated micromagnets¹¹, dynamic pumping of nuclear spins¹² or the addition of a third quantum dot¹⁷. Here we demonstrate a qubit that is a hybrid of spin and charge. It is simple, requiring neither nuclear-state preparation nor micromagnets. Unlike previous double-dot qubits, the hybrid qubit enables fast rotations about two axes of the Bloch sphere. We demonstrate full control on the Bloch sphere with π -rotation times of less than 100 picoseconds in two orthogonal directions, which is more than an order of magnitude faster than any other double-dot qubit. The speed arises from the qubit's charge-like characteristics, and its spin-like features result in resistance to decoherence over a wide range of gate voltages. We achieve full process tomography in our electrically controlled semiconductor quantum dot qubit, extracting high fidelities of 85 per cent for X rotations (transitions between qubit states) and 94 per cent for Z rotations (phase accumulation between qubit states).

As shown in Fig. 1a, the hybrid qubit^{20,21} is formed in a double quantum dot in a Si/SiGe heterostructure²², with the gate voltages tuned so that two electrons occupy the left dot and one electron occupies the right dot (the (2, 1) charge state). Changing the voltage on gate L to make the energy difference ε between the quantum dots more positive favours a transition to the (1, 2) charge state. By changing ε adiabatically from positive to negative, we can track the qubit state $|0\rangle \equiv |\downarrow\rangle|S\rangle$, where S denotes a singlet state in the right dot, from right to left, as shown by the dark blue line in Fig. 1b. Similarly, the state $|1\rangle \equiv 1/\sqrt{3}(|\downarrow\rangle|T_0\rangle + \sqrt{2/3}|\uparrow\rangle|T_-\rangle)$, where T_0 and T_- are two of the triplet states in the right dot, is shown by the magenta line in Fig. 1b.

The presence of the third electron in the pair of dots means that the states of the hybrid qubit are not purely singlet or triplet states, and we can therefore use fast electric field techniques to rotate the qubit about any axis of interest. In contrast, a two-electron singlet–triplet qubit requires slower, magnetic manipulation about one of the primary Bloch sphere axes. In addition, because the qubit is in only two dots rather than the three dots required for the exchange-only qubit^{16,17,23}, the detuning ε is the only control parameter. The gate voltage space for controlling the qubit is thus one dimensional, making it straightforward to find ‘sweet spots’ where the decoherence effects of charge fluctuations on the qubit are greatly reduced. One of these sweet spots, at $\varepsilon = 0$, is the same as is found in a conventional charge qubit. The main difference between the hybrid qubit and a pure charge qubit is the presence of a very wide sweet spot at positive detuning $\varepsilon \gg 0$. In this regime, the detuning of the dot suppresses the exchange interaction, making the energy levels parallel as a

function of detuning, and providing strong protection against charge noise.

We first discuss data that demonstrates rotations of the qubit about two axes on the Bloch sphere. As shown schematically in Fig. 1c, X rotations on the Bloch sphere correspond to oscillations between qubit states $|0\rangle$ and $|1\rangle$. To demonstrate such oscillations, we first prepare the qubit in state $|0\rangle$, by waiting for initialization at $\varepsilon = \varepsilon_r$ (Fig. 1b). Changing the detuning abruptly to ε_x results in a Hamiltonian $H = \mathcal{A}_1\sigma_x$, where σ_x is the usual Pauli matrix and the tunnel coupling \mathcal{A}_1 characterizes the strength of the anticrossing between $|0\rangle$ and $|1\rangle$. Such a Hamiltonian is expected to result in oscillations between states $|0\rangle$ and $|1\rangle$ at the Larmor frequency $2\mathcal{A}_1/\hbar$, where \hbar is Planck's constant. The resulting final state is measured by changing the detuning to ε_r : at this detuning, the charge-sensing quantum point contact is used to determine whether the charge state is (2, 1), corresponding to state $|0\rangle$, or (1, 2) corresponding to state $|1\rangle$ (see Methods Summary and Supplementary Information for details of the measurement procedure).

Figure 1e, f shows that this procedure results in rotations about the X axis of the Bloch sphere. In Fig. 1e we plot the probability P_0 of observing state $|0\rangle$ as a function of the pulse duration t_p and the gate voltage V_L , the latter of which determines ε . The path through Fig. 1e that corresponds to $\varepsilon = 0$ is curved (Fig. 1e, dashed curve), because of frequency-dependent attenuation in the microwave coaxial cable²⁴. Figure 1f shows a line cut through the path corresponding to $\varepsilon = 0$, revealing periodic oscillations in P_0 at a frequency of 5.2 GHz, corresponding to $\mathcal{A}_1/\hbar \approx 2.6$ GHz. The visibility of this oscillation is larger than 0.8, and we estimate a lower bound for the coherence time of $T_2^* \approx 2$ ns, by performing an exponential fit to the decay of the first 3 ns of the data. This X rotation alone is similar to a charge qubit rotation, and the coherence time is consistent with such an interpretation^{24–26}. The most important parameter describing a quantum gate is the process fidelity; we return to this quantity below after demonstrating Z rotations, because two rotation axes are required for process tomography.

Z rotations of the qubit, shown schematically in Fig. 1d as the orange line (Ramsey) about the equator of the Bloch sphere, can be performed by abruptly changing the detuning to a large, positive value, $\varepsilon_z > 100 \mu\text{eV}$. The states $|0\rangle$ and $|1\rangle$ have the same dependence on ε in this detuning range, differing in energy by a fixed value δE and creating a wide sweet spot that enables a controlled evolution of phase at a rate that is insensitive to fluctuations in ε arising, for example, from charge noise. Experimentally, the qubit is first prepared in state $|-Y\rangle \equiv 1/\sqrt{2}(|0\rangle - i|1\rangle)$, by initializing to state $|0\rangle$ and performing an $X(\pi/2)$ rotation, as described above. The resulting state can be rotated about the Z axis by setting the detuning equal to ε_z for a time t_p . We then measure the probability P_r of being in $|Y\rangle \equiv 1/\sqrt{2}(|0\rangle + i|1\rangle)$, by applying a second $X(\pi/2)$ rotation to rotate this state into state $|0\rangle$ and measuring the charge state at the readout position ε_r (see the orange inset to Fig. 1b). Figure 1g, h shows the resulting quantum oscillations of the qubit state around the Z axis of the Bloch sphere. The oscillations have a visibility larger than 0.85 and reveal a coherence time of $T_2^* \approx 10$ ns, which is much longer than

¹Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. ²Sandia National Laboratories, Albuquerque, New Mexico 87185, USA. ³Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

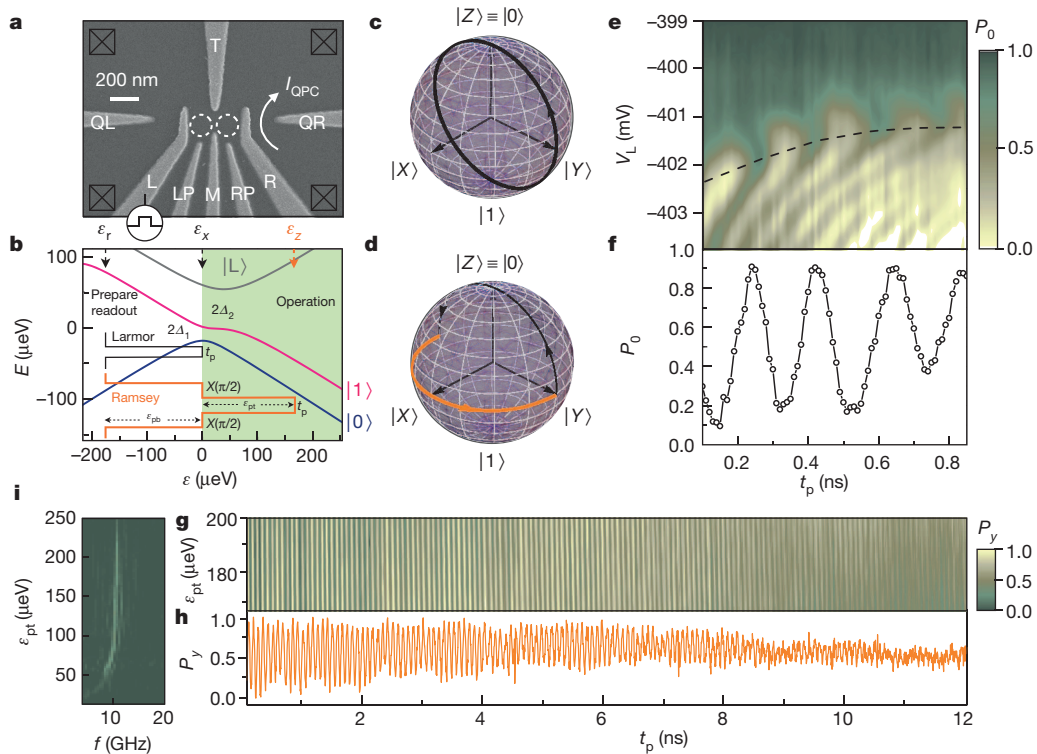


Figure 1 | Si/SiGe hybrid qubit device, energy levels and measurement of quantum oscillations. **a**, Scanning electron microscope image of a device lithographically identical to the one used in the experiment, with the locations of the double dot shown by white dashed circles. The current through the quantum point contact I_{QPC} is used for charge sensing via a measurement of its change in the presence of manipulation voltage pulses applied to gate L. The voltages on gates QL, QR, LP, M, RP and R are used to determine the overall operating point. **b**, Diagram of the calculated energy levels E versus detuning ε , including the ground states of the (2, 1) and (1, 2) charge configuration, the singlet–triplet splitting δE and the first excited state of the (1, 2) charge configuration, with Hamiltonian parameters determined as described in Supplementary Information. The resulting blue, magenta and grey solid curves show logical states $|0\rangle$ and $|1\rangle$ and the primary leakage state $|L\rangle$. Inset lines show pulse sequences used for implementation and measurement of rotations about the X axis (black) and Z axis (orange). **c**, **d**, Schematics of the evolution of the

Bloch vector during pulse sequences for X-axis (**c**) and Z-axis (**d**) rotations. **e**, Probability P_0 of the state to be $|0\rangle$, measured for the X-rotation pulse sequence as a function of the voltage V_L and pulse duration t_p . Dashed curve shows a path along $\varepsilon \approx 0$ μeV. **f**, Line cut of P_0 along the dashed curve in **e**, as a function of t_p , starting at $V_L \approx -402$ mV with pulse amplitude $\varepsilon_{\text{pb}} \approx 160$ μeV, showing ~ 5.2 GHz quantum oscillations with coherence time $T_2^* \approx 2$ ns. **g**, Probability P_y of the state being in $|Y\rangle \equiv 1/\sqrt{2}(|0\rangle + i|1\rangle)$, measured for the Z-rotation sequence as a function of pulse amplitude ε_{pt} and pulse duration t_p of the top pulse. Note the numerous parallel fringes, demonstrating that the frequency of Z rotation does not depend on ε_{pt} . **h**, Line cut of P_y as a function of t_p at $\varepsilon_{\text{pt}} \approx 180$ μeV, showing ~ 11.5 GHz oscillations with coherence time $T_2^* \approx 10$ ns. **i**, Fourier transform of Ramsey fringe data similar to that in **g**. The nearly vertical line at a frequency $f \approx 11$ –12 GHz corresponds to the Z-rotation frequency, indicating that the sweet spot extends for at least 150 μeV in the upper two-thirds of the plot.

the $Z(\pi/2)$ manipulation time of approximately 25 ps. As is clear from the parallel fringes in Fig. 1g, the oscillation frequency does not depend on ε , resulting in the long T_2^* . As shown in Fig. 1i, we can quantify the width in ε of the sweet spot by performing a Fourier transform of similarly acquired Ramsey fringe data. The main peak in the transform corresponds to the Ramsey fringe frequency, which is nearly constant for $\varepsilon_{\text{pt}} > 100$ μeV, where ε_{pt} is the detuning energy at which the Ramsey phase accumulation occurs.

The initialization, measurement and two rotation gates just described enable full control of the qubit and provide the basis for tomographic characterization of the resulting qubit state. Figure 2a shows schematic diagrams of three pulse sequences that achieve this goal. For each of the three sequences, the qubit state is prepared in $|0\rangle$ at ε_r , after which the detuning is pulsed to ε_x to perform an X rotation. By varying the time spent at ε_x from 160 to 340 ps—times that correspond to X rotation angles of approximately π to 3π —the qubit at the end of the initialization sequence (‘init’ in Fig. 2a) can be set to a controlled and nearly arbitrary superposition of $|0\rangle$ and $|1\rangle$. The limitation on this superposition is set by the X gate fidelity, which we extract below. In each of the three diagrams, this superposition state then evolves under a Z gate for a time t_p , reaching nearly all the rest of the Bloch sphere, with limitations again set by the Z gate fidelity.

Tomographic measurement is then performed by rotating the states $|X\rangle \equiv 1/\sqrt{2}(|0\rangle + |1\rangle)$, $|Y\rangle \equiv 1/\sqrt{2}(|0\rangle + i|1\rangle)$ and $|Z\rangle \equiv |0\rangle$ into state

$|0\rangle$, through the use of either a single X gate or a combination of an X gate and a Z gate, as shown by the blue lines in Fig. 2a. The resulting probability P_0 is measured by pulsing the detuning back to ε_r . This initialization and measurement scheme provides universal control of the hybrid qubit and allows us to reconstruct fully the time evolution of the state vector.

As an example of control of the hybrid qubit, we perform Z rotations on the Bloch sphere starting with two different initial states. Figure 2b–d shows the probabilities P_x , P_y and P_z of measuring the system in the states $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ following a Z rotation with the initial state either close to $|0\rangle$ (red data points) or close to $|-Y\rangle$ (purple data points). In Fig. 2e, f we plot the Bloch vectors that are extracted from these three measurements for each time t_p . The results obtained are consistent with intuitive expectations: for input state $|0\rangle$, the time evolution under a Z rotation accumulates a trivial phase and the Bloch vector remains near the north pole of the Bloch sphere. For input state $|-Y\rangle$, P_x and P_y oscillate sinusoidally between nearly 0 and 1, whereas P_z remains roughly constant with magnitude near 1/2, as expected.

Because the energy level structure of the qubit has been characterized experimentally, and the most likely leakage states are known (Fig. 1b), the rotation gates can be modelled numerically. The simulations incorporate realistic pulses with rise times of approximately 80 ps. The results of the calculations, shown as the solid orange ($|0\rangle$ initialization) and solid purple ($|-Y\rangle$ initialization) curves in Fig. 2b–f, are in good agreement

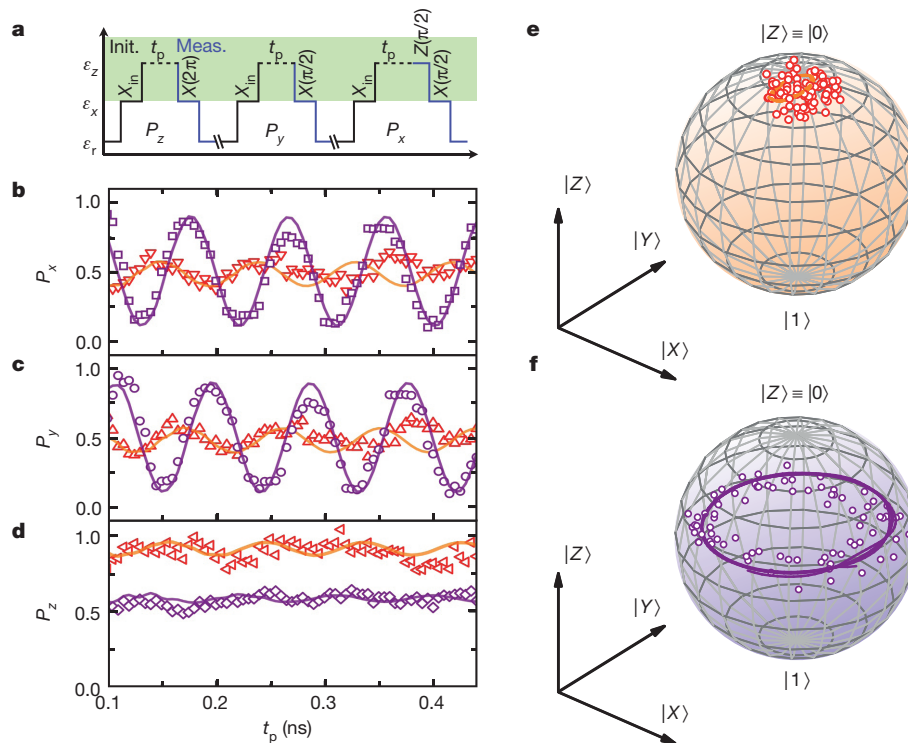


Figure 2 | State tomography and universal gate control of the hybrid qubit.

a, Schematic diagrams of the pulse sequences used to perform state tomography by measuring the Z-axis projection $P_z = P_0$, the Y-axis projection P_y , and the X-axis projection P_x of the state that is initialized through an X_{in} gate and rotated around the Z axis for an evolution time t_p . **b–d**, P_x , P_y and P_z as functions of t_p for states initialized near $|0\rangle$ (red symbols) and near $|1\rangle \equiv 1/\sqrt{2}(|0\rangle - i|1\rangle)$ (purple symbols). Solid curves with corresponding colours are calculated from a model using non-adiabatic pulses with 80 ps rise times and the energy level diagram of Fig. 1a, with best-fit parameters

$2A_1/h = 5.2$ GHz, $2A_2/h = 14.8$ GHz and $\delta E/h = 12.12$ GHz. The calculations neglect high-frequency dephasing but do include low-frequency fluctuations in the detuning ε (refs 23, 24 and Supplementary Information). **e**, Bloch sphere representation of the measured qubit state evolution under a Z gate that is initialized near $|0\rangle$. **f**, Representation with an input state near $|1\rangle$. Solid curves show the results of numerical simulations, as described above. The difference between **e** and **f** reflects the expected effect of Z rotations on the two different initial states.

with the experiment. As we describe in Supplementary Information, leakage into state $|L\rangle$ (shown in Fig. 1b) during these pulse sequences is 5% or less; in future work such leakage could be reduced further by appropriate pulse shaping. In Supplementary Information, we also report the results of an analogous state tomography of the qubit evolved under X rotations, using pulse sequences similar to those shown in Fig. 2a.

We now present quantum process tomography of the hybrid qubit. For a single qubit, the process matrix representation of any output state $\mathcal{E}(\rho)$ for a given input state ρ resulting from a given quantum process can be written as^{27,28}

$$\mathcal{E}(\rho) = \sum_{m,n=1}^4 \tilde{E}_m \rho \tilde{E}_n^\dagger \chi_{mn}$$

where the operators \tilde{E}_m form a basis in the space of 2×2 matrices, χ is the process matrix and a dagger denotes adjoint. To characterize both rotation axes, here we perform two sets of quantum process tomography: we consider both Z and X rotations, and for each we characterize rotations by an angle 2π (nominally equivalent to zero rotation or the identity gate), $\pi/2$ or π . To determine χ_{mn} for each process, we prepare four linearly independent input and output states using the manipulation approach presented above and in Fig. 2a. The process matrix is then obtained using maximum-likelihood estimation^{27,28}.

Figure 3 shows the results of this procedure; it reports in the Pauli basis $\{I, \sigma_x, \sigma_y, \sigma_z\}$ the real and imaginary parts of χ for both X and Z rotations of magnitude 2π , $\pi/2$ or π . For each process, we find reasonably good agreement between the estimated process matrix χ and the ideal process matrix χ_{ideal} . The process fidelity is $F_p = \text{Tr}(\chi_{ideal}\chi)$, yielding $F_p = 0.85 \pm 0.06$, 0.89 ± 0.05 and 0.88 ± 0.05 for the $X(2\pi)$, $X(\pi/2)$ and $X(\pi)$ processes, respectively, and 0.94 ± 0.02 , 0.96 ± 0.02 and 0.95 ± 0.02 for the

$Z(2\pi)$, $Z(\pi/2)$ and $Z(\pi)$ processes. The statistical uncertainty is estimated using 30 different sets of input and output states for each process (see Supplementary Information for all the fidelity results for each set of input and output states).

The qubit studied here was formed in natural Si, in which the low density of nuclear spins has sufficient spin protection that spin dephasing does not limit coherence. Furthermore, the absence of piezoelectric coupling between electrons and phonons eliminates a dephasing channel that would be important in polar semiconductors²⁹. With gate fidelities between 85% and 95%, the highest so far reported in an electrically gated semiconductor quantum dot qubit, the quantum dot hybrid qubit offers a promising combination of characteristics that in the past were found separately in qubits based on either charge or spin degrees of freedom: a good ratio of manipulation time to coherence time, together with fast overall operation and the ability to control the qubit fully using a single control parameter. It is important to emphasize the need for fast qubits: although the scaling associated with quantum computation is extremely promising for algorithms such as Shor's factoring algorithm and Grover's search algorithm, for the advantages to be realizable on time-scales relevant to humans, qubit operations must still be fast¹⁹. The fidelities of the X and Z gates reported here appear to be limited at present by the time spent in the regime near $\varepsilon = 0$, which is used for X rotations. In the future it may be possible to use a.c. driving of the control parameter to perform X rotations²⁹, so that the regime in which the hybrid qubit is sensitive to charge noise may be avoided entirely. Even when the $\varepsilon = 0$ regime is used for pulsed gating, as here, the hybrid qubit as demonstrated offers a desirable combination of high speed, high fidelity and efficient control in a semiconductor quantum dot qubit.

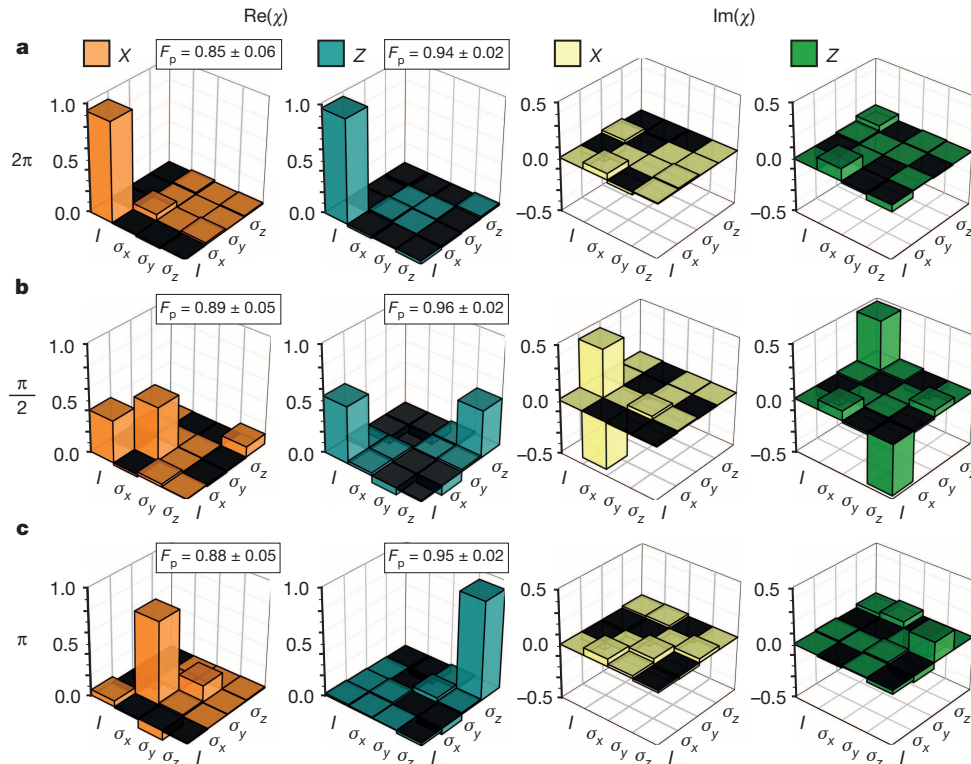


Figure 3 | Quantum process tomography of the hybrid qubit. Real and imaginary parts of the process matrix χ obtained by maximum-likelihood estimation for three X- and Z-rotation processes in the Pauli basis ($I, \sigma_x, \sigma_y, \sigma_z$). **a**, Rotation by 2π (identity); **b**, rotation by $\pi/2$; **c**, rotation by π . The measured process matrices are close to the targets for each operation; for example, the

METHODS SUMMARY

Measurement. The experiments are performed on a three-electron double quantum dot fabricated in a Si/SiGe heterostructure^{22,25,26,30} at base temperature (electron temperature, ~ 140 mK; ref. 30) in a dilution refrigerator. Fast voltage pulses are generated by Agilent 81134A pulse generator and are added to the dot-defining d.c. voltage through a bias tee (Picosecond Pulselabs 5546-107) before being applied to gate L. The conductance change through the quantum point contact with and without the manipulation pulses, measured with a lock-in amplifier (EG&G model 7265), is used to determine the average charge occupation and is converted to the reported probabilities. To partially compensate pulse distortion due to frequency-dependent attenuation in coaxial cables, we developed a pulse correction scheme based on qubit rotation and measurement, which is applied to all the sequences shown in Fig. 2. In Supplementary Information, we present the details of the measurement technique, the probability normalization and the pulse correction scheme.

Theory. Numerical simulations of the experiment were performed on the basis of the energy level diagram in Fig. 1b. Parameters in the simulation were extracted from experiment, as discussed in Supplementary Information. We model the dynamical evolution of the density matrix of the system as a function of detuning and pulse duration t_p using a master equation $\dot{\rho} = -(i/\hbar)[H, \rho]$ with the Hamiltonian H written in a basis of the position eigenstates^{25,26}. The (1, 2) charge occupation probability is extracted at the end of the pulse and is averaged over 2 ns in the measurement stage of the pulse. Low-frequency fluctuations in the detuning ε are incorporated by convolving the simulation result^{24,25} with a Gaussian distribution in ε of width $\sigma_\varepsilon = 5$ μ eV.

Received 15 January; accepted 28 April 2014.

- Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. K. Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79**, 1217–1265 (2007).
- Zwanenburg, F. A. et al. Silicon quantum electronics. *Rev. Mod. Phys.* **85**, 961–1019 (2013).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Hayashi, T., Fujisawa, T., Cheong, H. D., Jeong, Y. H. & Hirayama, Y. Coherent manipulation of electronic states in a double quantum dot. *Phys. Rev. Lett.* **91**, 226804 (2003).
- Petta, J. R., Johnson, A. C., Marcus, C. M., Hanson, M. P. & Gossard, A. C. Manipulation of a single charge in a double quantum dot. *Phys. Rev. Lett.* **93**, 186802 (2004).

target for an X($\pi/2$) rotation is 1/2 for the (I, I) and (σ_x, σ_x) components, $i/2$ and $-i/2$ for the (I, σ_x) and (σ_x, I) components, respectively, and zero for all others. For each process, the average value and uncertainty of the process fidelity $F_p = \text{Tr}(\chi_{\text{ideal}}\chi)$ is obtained from χ for 30 distinct sets of input and output states (Supplementary Information).

- Elzerman, J. M. et al. Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).
- Petta, J. R. et al. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
- Koppens, F. H. L. et al. Driven coherent oscillations of a single electron spin in a quantum dot. *Nature* **442**, 766–771 (2006).
- Kouwenhoven, L. P., Elzerman, J. M., Hanson, R., van Beveren, L. H. W. & Vandersypen, L. M. K. Control and measurement of electron spins in semiconductor quantum dots. *Phys. Stat. Solidi B* **243**, 3682–3691 (2006).
- Taylor, J. M. et al. Relaxation, dephasing, and quantum control of electron spins in double quantum dots. *Phys. Rev. B* **76**, 035315 (2007).
- Pioro-Ladrière, M. et al. Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nature Phys.* **4**, 776–779 (2008).
- Foletti, S., Bluhm, H., Mahalu, D., Umansky, V. & Yacoby, A. Universal quantum control of two-electron spin quantum bits using dynamic nuclear polarization. *Nature Phys.* **5**, 903–908 (2009).
- Laird, E. A. et al. Coherent spin manipulation in an exchange-only qubit. *Phys. Rev. B* **82**, 075403 (2010).
- Gaudreau, L. et al. Coherent control of three-spin states in a triple quantum dot. *Nature Phys.* **8**, 54–58 (2012).
- Maune, B. M. et al. Coherent singlet-triplet oscillations in a silicon-based double quantum dot. *Nature* **481**, 344–347 (2012).
- Medford, J. et al. Quantum-dot-based resonant exchange qubit. *Phys. Rev. Lett.* **111**, 050501 (2013).
- Medford, J. et al. Self-consistent measurement and state tomography of an exchange-only spin qubit. *Nature Nanotechnol.* **8**, 654–659 (2013).
- Dial, O. E. et al. Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit. *Phys. Rev. Lett.* **110**, 146804 (2013).
- Van Meter, R., Itoh, K. M. & Ladd, T. D. in *Controllable Quantum States: Mesoscopic Superconductivity & Spintronics (MS+S2006)*, Proc. Int. Symp. (eds Takayanagi, H., Nitta, J. & Nakano, H.) 183–188 (World Scientific, 2008); preprint at <http://arxiv.org/abs/quant-ph/0507023> (2006).
- Shi, Z. et al. Fast hybrid silicon double quantum dot qubit. *Phys. Rev. Lett.* **108**, 140503 (2012).
- Koh, T. S., Gamble, J. K., Friesen, M., Eriksson, M. A. & Coppersmith, S. N. Pulse-gated quantum dot hybrid qubit. *Phys. Rev. Lett.* **109**, 250503 (2012).
- Thalakulam, M. et al. Fast tunnel rates in Si/SiGe one-electron single and double quantum dots. *Appl. Phys. Lett.* **96**, 183104 (2010).
- DiVincenzo, D. P., Bacon, D., Kempe, J., Burkard, G. & Whaley, K. B. Universal quantum computation with the exchange interaction. *Nature* **408**, 339–342 (2000).
- Petersson, K. D., Petta, J. R., Lu, H. & Gossard, A. C. Quantum coherence in a one-electron semiconductor charge qubit. *Phys. Rev. Lett.* **105**, 246804 (2010).

25. Shi, Z. *et al.* Coherent quantum oscillations and echo measurements of a Si charge qubit. *Phys. Rev. B* **88**, 075416 (2013).
26. Shi, Z. *et al.* Fast coherent manipulation of three-electron states in a double quantum dot. *Nature Commun.* **5**, 3020 (2014).
27. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 389–393 (Cambridge Univ. Press, 2000).
28. Chow, J. M. *et al.* Randomized benchmarking and process tomography for gate errors in a solid-state qubit. *Phys. Rev. Lett.* **102**, 090502 (2009).
29. Koh, T. S., Coppersmith, S. N. & Friesen, M. High-fidelity gates in quantum dot spin qubits. *Proc. Natl Acad. Sci. USA* **110**, 19695–19700 (2013).
30. Simmons, C. B. *et al.* Tunable spin loading and T_1 of a silicon spin qubit measured by single-shot readout. *Phys. Rev. Lett.* **106**, 156804 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported in part by ARO (W911NF-12-0607), the NSF (PHY-1104660) and by the Laboratory Directed Research and Development programme at Sandia National Laboratories. Sandia National Laboratories is a multi-programme laboratory managed and operated by Sandia Corporation, a wholly

owned subsidiary of Lockheed Martin Corporation, for the US Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Development and maintenance of the growth facilities used for fabricating samples is supported by the US Department of Energy (DE-FG02-03ER46028). This research used US National Science Foundation-supported shared facilities at the University of Wisconsin-Madison. D.K. acknowledges conversations with X. Wu and K. Rudinger.

Author Contributions M.A.E. and S.N.C. had the idea for the experiment. D.K. developed pulse sequences for qubit operation and tomography, performed electrical measurements and numerical simulations with the aid of Z.S., and analysed the data with M.A.E. and S.N.C. C.B.S. fabricated the quantum dot device. J.R.P. and D.R.W. developed hardware and software for the measurements. T.S.K., J.K.G. and M.F. helped with the theoretical analysis. D.E.S. and M.G.L. prepared the Si/SiGe heterostructure. D.K., S.N.C. and M.A.E. wrote the manuscript with the contributions of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.E. (maeriksson@wisc.edu).

Abrupt pre-Bølling-Allerød warming and circulation changes in the deep ocean

Nivedita Thiagarajan¹, Adam V. Subhas¹, John R. Southon², John M. Eiler¹ & Jess F. Adkins¹

Several large and rapid changes in atmospheric temperature and the partial pressure of carbon dioxide in the atmosphere¹—probably linked to changes in deep ocean circulation²—occurred during the last deglaciation. The abrupt temperature rise in the Northern Hemisphere and the restart of the Atlantic meridional overturning circulation at the start of the Bølling-Allerød interstadial, 14,700 years ago, are among the most dramatic deglacial events³, but their underlying physical causes are not known. Here we show that the release of heat from warm waters in the deep North Atlantic Ocean probably triggered the Bølling-Allerød warming and reinvigoration of the Atlantic meridional overturning circulation. Our results are based on coupled radiocarbon and uranium-series dates, along with clumped isotope temperature estimates, from water column profiles of fossil deep-sea corals in a limited area of the western North Atlantic. We find that during Heinrich stadial 1 (the cool period immediately before the Bølling-Allerød interstadial), the deep ocean was about three degrees Celsius warmer than shallower waters above. This reversal of the ocean's usual thermal stratification pre-dates the Bølling-Allerød warming and must have been associated with increased salinity at depth to preserve the static stability of the water column. The depleted radiocarbon content of the warm and salty water mass implies a long-term disconnect from rapid surface exchanges, and, although uncertainties remain, is most consistent with a Southern Ocean source. The Heinrich stadial 1 ocean profile is distinct from the modern water column, that for the Last Glacial Maximum and that for the Younger Dryas, suggesting that the patterns we observe are a unique feature of the deglacial climate system. Our observations indicate that the deep ocean influenced dramatic Northern Hemisphere warming by storing heat at depth that preconditioned the system for a subsequent abrupt overturning event during the Bølling-Allerød interstadial.

Understanding the cause of abrupt terminations of glacial periods is a central question in palaeoclimate as these rapid warmings characterize the 'sawtooth' nature of glacial cycles⁴. The last deglaciation started ~18 kyr ago in the Southern Hemisphere with an increase in temperature¹ and greenhouse gas concentration¹ and a retreat of sea ice⁵. For over 3 kyr Antarctica warmed while Greenland and the North Atlantic remained cold. This period, known as Heinrich stadial 1 (HS1), ended abruptly 14.7 kyr ago at the start of the Bølling-Allerød, when sea ice retreated in the north⁵, the Atlantic meridional overturning circulation (AMOC) restarted² and the Northern hemisphere dramatically warmed¹, with a corresponding return to cold conditions in the south¹. Explanations for the Bølling-Allerød shift have focused on the AMOC restart and include a variety of triggers^{6–9}. Here we use coupled U-series and ¹⁴C ages and clumped isotope (Δ_{47} ; Methods) temperature measurements on deep-sea corals from the New England seamounts to reconstruct ocean temperatures and circulation along a depth transect (Methods). We find that the deep ocean at our site was characterized by warmer waters underneath colder waters ~800 yr before the beginning of the Bølling-Allerød. This unusual scenario implies that salt stratification of the water column stored potential energy at depth, giving the deep ocean a fundamental role in the restart of the AMOC at the beginning of the Bølling-Allerød.

Figure 1 shows new and previously published^{10,11} $\Delta^{14}\text{C}$ values for the Younger Dryas and two time slices during HS1. Prior to 15.6 kyr ago, the $\Delta^{14}\text{C}$ of the water column was offset from the contemporaneous atmosphere by 110–195‰. This value is much larger than the modern North Atlantic offset of ~70‰, implying that the HS1 ocean was less well ventilated (older) than the modern ocean. At that time, $\Delta^{14}\text{C}$ decreased from top to bottom in the profile. At 15.6 kyr ago there was a sudden, large shift to even more ¹⁴C-depleted waters at intermediate depths. This 'mid-15-kyr' event is observed within the skeletons of two separate corals, JFA 24.8 and JFA 24.19, both found at 1,784 m. As previously reported, this large shift, contemporaneous with a constant $\Delta^{14}\text{C}$ at greater depths (2,000–2,400 m), is consistent with a lateral influx of very ¹⁴C-depleted, southern-sourced intermediate waters^{10,11}, implying a water mass boundary at depth between rapidly changing upper waters and less variant deep waters.

The timing of this change in $\Delta^{14}\text{C}$ is surprising because these mid-15-kyr changes do not appear to correlate in time with either the air temperature records from Greenland ice cores or the largest changes in atmospheric gas concentration¹. However, mid-HS1 variability has previously been observed and was first described in the character and source region of icebergs reaching the Iberian margin¹² (called the HS1a–HS1b transition), and pre-Bølling-Allerød changes have been seen in other marine records (Methods). Here we propose a scenario where the mid-15-kyr changes are unique to the intermediate and upper portions of the deep ocean.

The most striking new feature of the deglacial water column comes from a subset of the corals in Fig. 1a–c that we analysed for Δ_{47} -based temperatures (Figs 1d–f and 2). These corals reveal a large (5 °C) and abrupt warming during HS1 in the North Atlantic that is almost contemporaneous with the shift to even more ¹⁴C-depleted waters. These signals must arise from the movement of a front across our corals during their lifetime. By running numerous replicates, we were able to shrink the uncertainty in our temperature estimates and confirm that the deep warming is outside the limit of our analytical errors¹³ (Methods). This abrupt warming has an important depth structure, as documented by temperature profiles before and after the $\Delta^{14}\text{C}$ event (Fig. 1d–f). The pre-15.6-kyr profile is cooler than the modern one and is roughly isothermal, but the younger HS1 profile is cool at 1,381 m and warmer at greater depths. To maintain static stability, this temperature inversion requires that the water column also becomes saltier with depth, a feature that has been previously observed for the Last Glacial Maximum¹⁴ (LGM). If we assume constant potential density (a minimum definition of static stability) between 1,381 and 2,600 m, we calculate that the smallest salt difference needed to support the observed warm deep waters below shallower cooler waters is ~0.8 p.s.u., which is well within the bounds of the LGM ocean¹⁴.

The deglacial coevolution of temperature and $\Delta^{14}\text{C}$ at intermediate depths shows three vectors of ocean temperature and $\Delta^{14}\text{C}$ behaviour that occur over the course of HS1, one from 17 to 15.8 kyr ago, when the ocean became more depleted in $\Delta^{14}\text{C}$ (even relative to the modern atmosphere); one around 15.8 to 15.6 kyr ago, when waters deeper than

¹Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ²Earth System Science Department, 3200 Croul Hall, University of California, Irvine, California 92697–3100, USA.

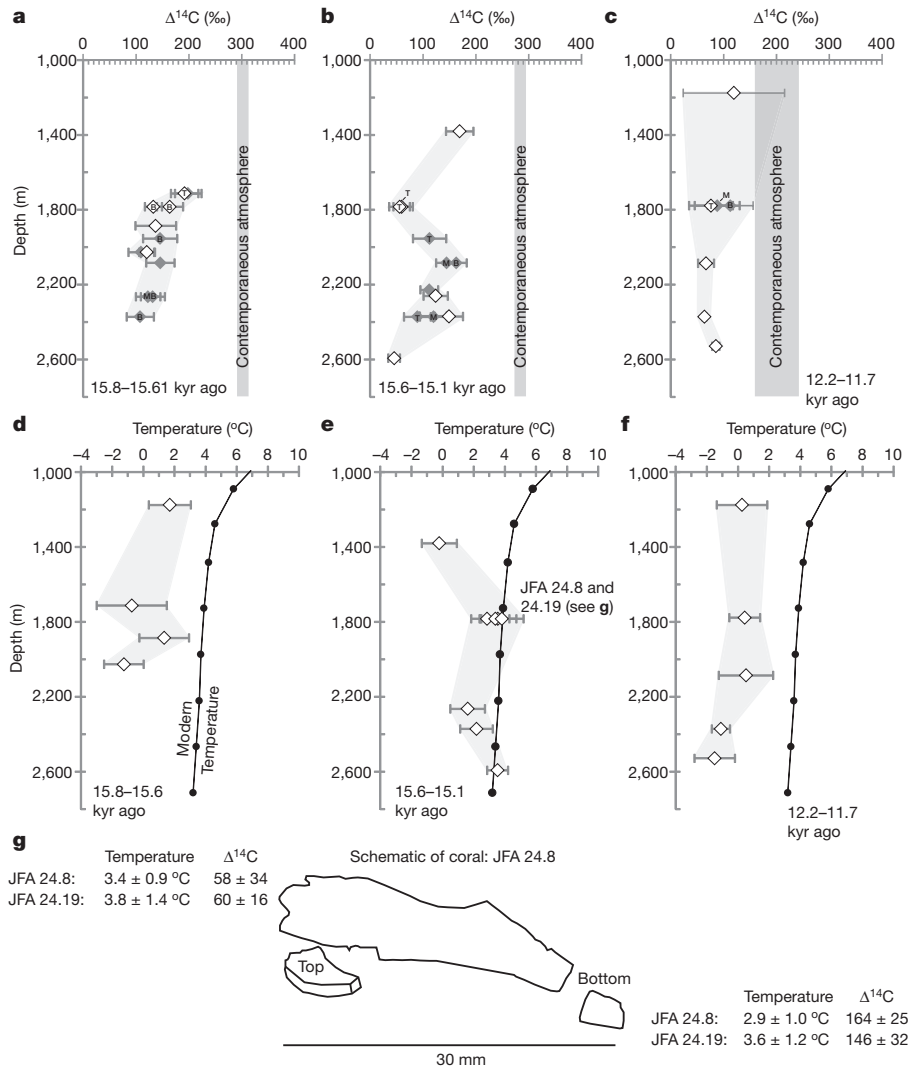


Figure 1 | Ocean profiles of $\Delta^{14}\text{C}$ and temperature during three different time intervals. **a–c**, $\Delta^{14}\text{C}$ profiles. Points outlined in black have also been measured for temperature. There is an abrupt shift in $\Delta^{14}\text{C}$ ($\Delta^{14}\text{C} = 1,000 \times (e^{-14C_{\text{age}}/8,033} / e^{-14C_{\text{atm}}/8,266} - 1)$), where $^{14}\text{C}_{\text{age}}$ and Cal_{age} are respectively the radiocarbon age and the calendar age of the coral) at mid-15-kyr. T, M and B refer to analyses made on the top, middle and bottom

of a coral sample, respectively. Symbols with no letters refer to the top of the coral. Error bars are 1 s.e.m. **d–f**, Temperature profiles. Unlike the $\Delta^{14}\text{C}$ profiles, the bottoms of JFA 24.8 and JFA 24.19 are on the 15.6–15.1-kyr panel (**e**). The value of n for each measurement is given in Source Data. **g**, Sketch of coral transects. $\Delta^{14}\text{C}$ shows an abrupt shift between the tops and bottoms of the two corals, but there is little change in temperature.

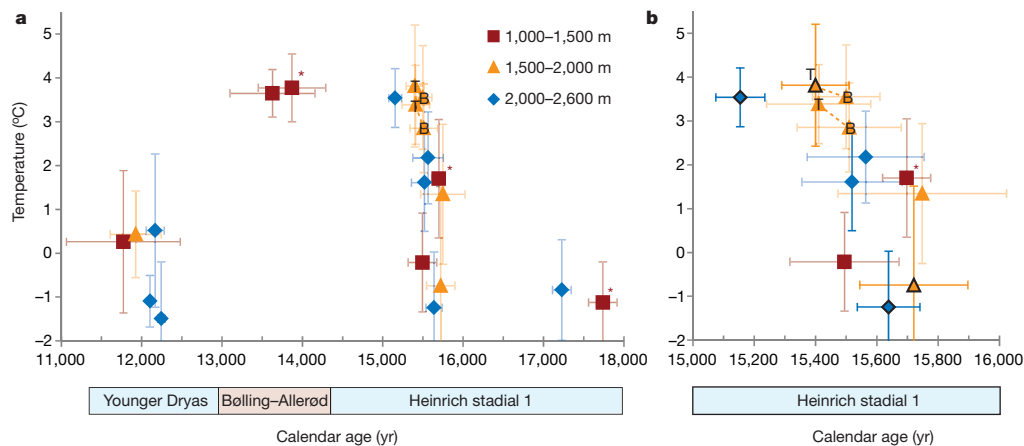


Figure 2 | Temperature versus calendar age based on clumped isotopes in deep-sea corals across the deglaciation. **a**, The dashed lines connect the tops and bottoms of JFA 24.8 and JFA 24.19. T and B refer to the top and bottom of a coral, respectively. Error bars are 1 s.e.m. Asterisks indicate corals that have either a high $\delta^{234}\text{U}_i$ (which is the initial $^{234}\text{U}/^{238}\text{U}$ isotope ratio at the time of

coral growth) or a $\Delta^{14}\text{C}$ greater than the atmosphere value. In both cases, this open-system behaviour changes the $\Delta^{14}\text{C}$ values but does not change calendar ages much on this plot. The value of n for each measurement is given in Source Data. **b**, The abrupt warming is encapsulated by the two oldest and youngest corals within this time frame, and these corals are outlined in black.

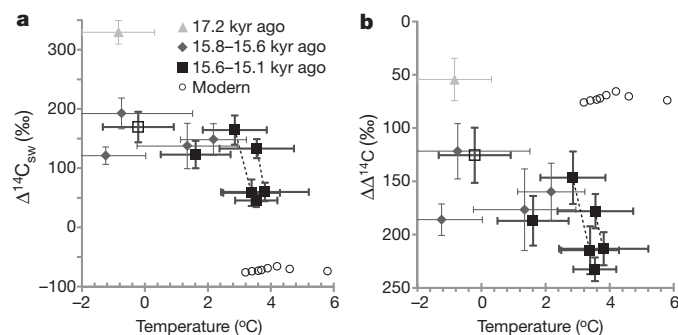


Figure 3 | $\Delta^{14}\text{C}$ -temperature cross-plot. **a**, Plot of $\Delta^{14}\text{C}$ versus temperature for the intermediate-depth waters of the deglacial North Atlantic. The open square is the shallowest HS1 coral, from 1,381 m, and does not warm. The dashed lines connect the tops and bottoms of JFA 24.8 and JFA 24.19. Modern $\Delta^{14}\text{C}$ and temperature measurements for our site are from the GEOSECS database. The value of n for each measurement is given in Source Data. **b**, Offset of reconstructed deep-sea coral $\Delta^{14}\text{C}$ from contemporaneous atmosphere as determined by the Hulu record³⁰ versus temperature for intermediate depths. Error bars are 1 s.e.m.

1,600 m warmed; and one from 15.6 to 15.1 kyr ago, when those same deeper waters became even more depleted in $\Delta^{14}\text{C}$ (Fig. 3). The temperature and $\Delta^{14}\text{C}$ transitions are not simultaneous (Fig. 1g); both the tops and the bottoms of corals JFA 24.8 and JFA 24.19 are already warm, and the $\Delta^{14}\text{C}$ switch occurs during their lifetime. This offset, although unusual, has several possible explanations. During the deglaciation, the ocean-atmosphere system was not at steady state, and it is therefore possible that multiple reservoirs with different heat and radiocarbon contents influenced our site during HS1. Even for a single water mass, temperature and carbon isotopes have different timescales for exchange at their outcrop region, so they may have been decoupled from each other for ~ 100 yr during deglaciation. Regardless of the cause of the offset, the warm, $\Delta^{14}\text{C}$ -depleted water that existed by 15.4 ± 0.2 kyr ago, below a colder upper layer, is a uniquely deglacial feature, with no equivalent in either the modern or the LGM oceans.

The observation of warming at intermediate depths in the Atlantic during HS1 has been controversial. Several regional benthic foraminifera records show depleted $\delta^{18}\text{O}$ values that were initially interpreted as a signature of brine rejection in the Nordic Seas¹⁵, but other temperature records attribute the same signal to intermediate or deep ocean warming^{16,17}. On the basis of our clumped isotope data, at least $\sim 1\%$ of the benthic $\delta^{18}\text{O}$ depletion seen in these records is due to warming. Several modelling studies^{7,16,18} also show intermediate or deep North Atlantic warming during deglaciation, due to diffusion of tropical ocean heat across the main thermocline during times of slow North Atlantic overturning. If the downward diffusion of heat were local, this scenario would predict monotonically decreasing temperature profiles with depth¹⁶, but we see cold shallow waters above warm deeper waters at the mid-15kyr event. Given the abrupt nature of the warming in our corals, the signal is most probably that of heat that is accumulated elsewhere and quickly brought to $\sim 2,000$ m water depth by movement of a deep front.

Other possible sources of heat to the intermediate or deep ocean include geothermal heating¹⁹ and warming of the outcrop region for isopycnals that occupied our site: either could lead to the temperature inversion we observe. Several water masses have been found during Heinrich events to have either warmed or exhibited depleted $\Delta^{14}\text{C}$ values, and, although none of these show both signatures simultaneously (Methods), an influx of southern-sourced water to our site could explain our results (Extended Data Figs 6 and 7), because inefficient carbon isotope exchange in the outcrop regions tends to produce low initial $\Delta^{14}\text{C}$ values. Heating at the deep-water formation zones in the Southern Hemisphere could be transferred to near our site by the deep circulation. Then a pre-Bølling-Allerød shift in deep-water structure (potentially caused by winds or buoyancy

fluxes in the Southern Ocean, or both) could cause this warm front to cross our site from deeper waters to shallower.

Although the geographic origin of the warm deep waters is an important issue, the most significant feature of our new results is that the presence of warm waters below colder waters, supported by salt stratification, is a natural way to build capacitance in the climate system. Warming of the deep ocean can explain the polar records of deglaciation and helps us understand how the ocean moved between the stable glacial and interglacial states of deep circulation.

A new compilation of Atlantic benthic $\delta^{13}\text{C}$ over the deglaciation suggests that the transition from HS1 to the Bølling-Allerød is the moment where the ocean switched between these two stable circulation states (Fig. 4 and Extended Data Fig. 8). Sections of $\delta^{13}\text{C}$ show a vertically stratified ocean at the LGM with large $\delta^{13}\text{C}$ differences between southern and northern endmembers. In the early deglaciation, the upper waters changed more than the deepest waters (Extended Data Fig. 8), reflecting a more dynamic interplay between northern- and southern-sourced intermediate and upper deep waters. This situation changed drastically at the HS1/Bølling-Allerød transition, when the deep southern-sourced²⁰ water mass became much less depleted and volumetrically less important. Pa/Th records^{2,21,22} exhibit similar behaviour during the early deglaciation, with intermediate depth water masses showing a more vigorous circulation than the more stable deep-ocean records^{2,22}. This was followed by the HS1/Bølling-Allerød transition, when, according to deep-ocean Pa/Th, the AMOC was reinvigorated. This reinvigoration continued to its modern strength with a small decline during the Younger Dryas.

We suggest that the transition between LGM and modern circulation patterns was facilitated by instabilities in the deep water column. A water column with cold, fresh water on top of warm, salty water, as observed after the mid-15-kyr event, can be susceptible to a finite-amplitude instability, where water parcels rapidly exchange positions vertically to release potential energy stored in as heat in the deep ocean²³. This condition is analogous to convectively available potential energy in the atmosphere²³. In the ocean it is thermobaricity²⁴ that leads to the instability, and its effects are seen in deep-water formation regions today²⁵ as well as in

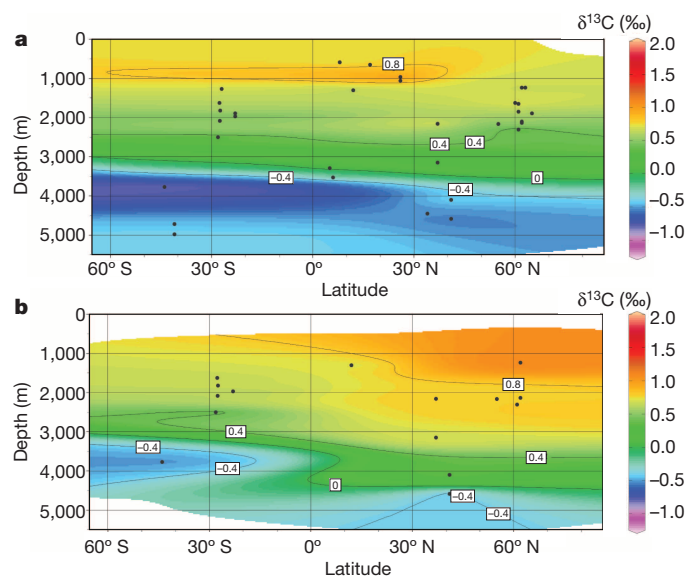


Figure 4 | Benthic sections of $\delta^{13}\text{C}$ during late HS1 and the Bølling-Allerød. Well-dated, high-resolution cores from the Atlantic Ocean were chosen from the end of HS1 (14.5–15.7 kyr ago) (**a**) and during the Bølling-Allerød (13–14.5 kyr ago) (**b**). Black dots indicate the latitudes and depths of cores used to make the sections. These time slices come from a larger set across the whole termination (Extended Data Fig. 8), and we display them because they show the largest switch from southern-source-dominated waters in the abyssal Atlantic at the LGM to a more ‘interleaved’ pattern in the modern. The Bølling-Allerød marks this abrupt shift in deep tracer distribution.

large eddy simulation models of ocean convection²⁶. In this scenario, the 'trigger' for the Bølling–Allerød AMOC restart and abrupt warming was not the removal of fresh water from the surface of the North Atlantic²⁷, but was instead the release of deep-ocean warmth that built up over time. A trigger for the release can always be found by the climate system (breaking internal waves that initiate vertical water movement, and small changes in the buoyancy budget of the surface ocean are two examples) but it is less fundamental to the physics than the build-up of the potential energy itself. Rather than being merely a response to fresh water forcing at the surface²⁸, the deep ocean circulation has an important role in the forcing of the dramatic Northern Hemisphere warming at the start of the Bølling–Allerød and in the bipolar nature of deglaciations generally. On this view, the Bølling–Allerød represents a fundamental change in the density structure of the deep ocean, when the thermobaric instability of the whole deep water column allowed the system to 'punch through' the LGM stratification and move towards the modern deep Atlantic structure of warm, salty North Atlantic deep water and cold, fresh Antarctic bottom water (Fig. 4). Whereas the Younger Dryas water column in the North Atlantic is often thought of as a return to Heinrich-like water masses, we show that it is structured very differently than the HS1 water column in temperature, $\delta^{13}\text{C}$ and $\Delta^{14}\text{C}$. Mid-depth warming in the North Atlantic has also been observed at many Heinrich events¹⁶, but only HS1 led to a deglaciation. Future work will have to establish whether salt stratification accompanied these previous temperature changes.

METHODS SUMMARY

Samples examined in this study were obtained from the California Institute of Technology deep-sea coral fossil collection. The deep-sea coral species *Desmophyllum dianthus* from the New England (34–40° N, 60–68° W) and Corner Rise seamounts (34–36° N, 47–53° W) (Extended Data Fig. 1) were collected in 2003–2005 using the deep submergence vehicles *Alvin* and *Hercules*. Collection by remotely operated underwater vehicle and submarine ensured that the depth of each coral was precisely known and that corals were collected near growth positions. All samples in this study were cleaned and analysed for U/Th ages, ^{14}C ages and Δ_{47} temperatures (Supplementary Information).

All samples analysed for ^{14}C and U/Th were physically cleaned and chemically cleaned following previously established procedures^{11,29} (Methods). Radiocarbon ages were measured at the UC-Irvine Keck-CCAMS facility on an accelerator mass spectrometer. Uranium-series ages were run on a multi-collector inductively coupled plasma mass spectrometer at Caltech. All samples were bracketed by an instrumental standard of known isotopic composition to correct for mass bias and mass drift.

Clumped isotope measurements were made on the deep-sea corals at the California Institute of Technology following previously established procedures¹³. Deep-sea corals, unlike modern corals, have a Fe–Mn crust, and a cleaning study determined that physical cleaning alone gave accurate clumped isotope temperatures (Extended Data Fig. 2). Samples were analysed multiple times ($n = 3$ –20) to ensure low external standard errors.

Benthic $\delta^{13}\text{C}$ sections were made by compiling well-dated, high-resolution benthic $\delta^{13}\text{C}$ records (Methods).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 August 2013; accepted 7 May 2014.

- Petit, J.-R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- McManus, J., Francois, R., Gherardi, J., Keigwin, L. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Clark, P. U. *et al.* Global climate evolution during the last deglaciation. *Proc. Natl Acad. Sci. USA* **109**, E1134–E1142 (2012).
- Broecker, W. S. & van Donk, J. Insolation changes, ice volumes, and the O18 record in deep-sea cores. *Rev. Geophys.* **8**, 169–198 (1970).

- Denton, G. H., Alley, R. B., Comer, G. C. & Broecker, W. S. The role of seasonality in abrupt climate change. *Quat. Sci. Rev.* **24**, 1159–1182 (2005).
- Weaver, A. J., Saenko, O. A., Clark, P. U. & Mitrovica, J. X. Meltwater pulse 1A from Antarctica as a trigger of the Bølling–Allerød warm interval. *Science* **299**, 1709–1713 (2003).
- Liu, Z. *et al.* Transient simulation of last deglaciation with a new mechanism for Bølling–Allerød warming. *Science* **325**, 310–314 (2009).
- Rogerson, M. *et al.* Enhanced Mediterranean–Atlantic exchange during Atlantic freshening phases. *Geochim. Geophys. Geosyst.* **11**, Q08013 (2010).
- Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424**, 532–536 (2003).
- Adkins, J. F., Cheng, H., Boyle, E. A., Druffel, E. R. M. & Edwards, L. Deep-sea coral evidence for rapid change in ventilation of the deep North Atlantic 15,400 years ago. *Science* **280**, 725–728 (1998).
- Robinson, L. F. *et al.* Radiocarbon variability in the western North Atlantic during the last deglaciation. *Science* **310**, 1469–1473 (2005).
- Bard, E., Rostek, F., Turon, J.-L. & Gendreau, S. Hydrological impact of Heinrich events in the subtropical Northeast Atlantic. *Science* **289**, 1321–1324 (2000).
- Thiagarajan, N., Adkins, J. & Eiler, J. Carbonate clumped isotope thermometry of deep-sea corals and implications for vital effects. *Geochim. Cosmochim. Acta* **75**, 4416–4425 (2011).
- Adkins, J. F., McIntyre, K. & Schrag, D. P. The salinity, temperature, and $\delta^{18}\text{O}$ of the glacial deep ocean. *Science* **298**, 1769–1773 (2002).
- Dokken, T. & Jansen, E. Rapid changes in the mechanism of ocean convection during the last glacial period. *Nature* **401**, 458–461 (1999).
- Marcott, S. A. *et al.* Ice-shelf collapse from subsurface warming as a trigger for Heinrich events. *Proc. Natl Acad. Sci.* **108**, 13415–13419 (2011).
- Rasmussen, T. L., Thomsen, E., Labeyrie, L. & van Weering, T. C. E. Circulation changes in the Faeroe–Shetland Channel correlating with cold events during the last glacial period (58–10 ka). *Geology* **24**, 937–940 (1996).
- Shaffer, G., Olsen, S. M. & Bjerrum, C. J. Ocean subsurface warming as a mechanism for coupling Dansgaard-Oeschger climate cycles and ice-rafting events. *Geophys. Res. Lett.* **31**, L24202 (2004).
- Stein, C. A. & Stein, S. A model for the global variation in oceanic depth and heat-flow with lithospheric age. *Nature* **359**, 123–129 (1992).
- Kwon, E. Y. *et al.* North Atlantic ventilation of “southern-sourced” deep water in the glacial ocean. *Paleoceanography* **27**, PA2208 (2012).
- Gherardi, J.-M. *et al.* Evidence from the Northeastern Atlantic basin for variability in the rate of the meridional overturning circulation through the last deglaciation. *Earth Planet. Sci. Lett.* **240**, 710–723 (2005).
- Gherardi, J. *et al.* Glacial–interglacial circulation changes inferred from $^{231}\text{Pa}/^{230}\text{Th}$ sedimentary record in the North Atlantic region. *Paleoceanography* **24**, PA2204 (2009).
- Adkins, J. F., Ingersoll, A. P. & Pasquero, C. Rapid climate change and conditional instability of the glacial deep ocean from the thermobaric effect and geothermal heating. *Quat. Sci. Rev.* **24**, 581–594 (2005).
- McDougall, T. Thermobaricity, cabbelling, and water-mass conversion. *J. Geophys. Res.* **92**, 5448–5464 (1987).
- Denbo, D. W. & Skillingstad, E. D. An ocean large-eddy simulation model with application to deep convection in the Greenland Sea. *J. Geophys. Res.* **101**, 1095–1110 (1996).
- Akitomo, K. Open-ocean deep convection due to thermobaricity 1. Scaling argument. *J. Geophys. Res.* **104**, 5225–5234 (1999).
- Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
- Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
- Adkins, J. F. *et al.* Radiocarbon dating of deep-sea corals. *Radiocarbon* **44**, 567–580 (2002).
- Southon, J., Noronha, A. L., Cheng, H., Edwards, R. L. & Wang, Y. A high-resolution record of atmospheric ^{14}C based on Hulu Cave speleothem H82. *Quat. Sci. Rev.* **33**, 32–41 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. McManus and M. Miller for discussions. We also thank the captain and crew of the RV *Atlantis* cruise AT7-35 and the WHOI Deep Submergence *Alvin* and *ABE* groups.

Author Contributions N.T. and J.F.A. designed the study. N.T. collected the ^{14}C and Δ_{47} data and compiled the benthic $\delta^{13}\text{C}$ sections. J.R.S. facilitated and oversaw the ^{14}C measurements and J.M.E. facilitated and oversaw the Δ_{47} measurements. A.V.S. collected the U-series data. J.F.A. facilitated and oversaw the U-series measurements. N.T. and J.F.A. wrote the first draft of the manuscript. All authors contributed to the interpretation and preparation of the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.T. (nivedita@gps.caltech.edu).

New specimen of *Archaeopteryx* provides insights into the evolution of pennaceous feathers

Christian Foth^{1,2}, Helmut Tischlinger³ & Oliver W. M. Rauhut^{1,2}

Discoveries of bird-like theropod dinosaurs and basal avialans in recent decades have helped to put the iconic ‘Urvogel’ *Archaeopteryx* into context^{2–6} and have yielded important new data on the origin and early evolution of feathers⁷. However, the biological context under which pennaceous feathers evolved is still debated. Here we describe a new specimen of *Archaeopteryx* with extensive feather preservation, not only on the wings and tail, but also on the body and legs. The new specimen shows that the entire body was covered in pennaceous feathers, and that the hindlimbs had long, symmetrical feathers along the tibiotarsus but short feathers on the tarsometatarsus. Furthermore, the wing plumage demonstrates that several recent interpretations^{8,9} are problematic. An analysis of the phylogenetic distribution of pennaceous feathers on the tail, hindlimb and arms of advanced maniraptorans and basal avialans strongly indicates that these structures evolved in a functional context other than flight, most probably in relation to display, as suggested by some previous studies^{10–12}. Pennaceous feathers thus represented an exaptation and were later, in several lineages and following different patterns, recruited for aerodynamic functions. This indicates that the origin of flight in avialans was more complex than previously thought and might have involved several convergent achievements of aerial abilities.

The iconic ‘Urvogel’ *Archaeopteryx* has long been the ‘yardstick’ for bird evolution¹. However, in the past 15 years our knowledge on early bird evolution has increased significantly, mainly because of the discovery of bird-like theropods and early avialans with feather preservation from the Middle–Late Jurassic and Early Cretaceous periods of China^{2–6}. This has provided a wealth of new data on the evolution of feathers and their functional context, which has challenged many traditional ideas^{3,7,13}. Although *Archaeopteryx* still plays a pivotal role in discussions of avialan origins and basal relationships^{5,6,14}, information on the plumage of this crucial taxon has so far mainly been limited to the London and Berlin specimens, with some additional information from the recently described Thermopolis specimen.

The new, excellently preserved 11th skeletal specimen of *Archaeopteryx* shows extensive feather preservation, not only along the wings and tail (as in the London, Berlin and Thermopolis specimens), but also in other parts of the body (Fig. 1), and thus helps to clarify controversial issues about the plumage of basal avialans. Most of the skeleton is preserved in articulation, but the skull is incomplete and removed from the neck, and the right pectoral girdle and wing have been displaced to lie below the left hindlimb; the left wing is missing (Fig. 1, Extended Data Figs 1 and 2 and Extended Data Table 1).

The plumage of the new specimen is preserved mainly as imprints, with a few patches showing possible preservation of organic material. Remains of the body plumage are preserved along the presacral axial skeleton and the pelvis (Fig. 1 and Extended Data Fig. 3). All body feathers show a high degree of overlapping, but because of excellent preservation the identification of feather structures is possible. The body plumage of *Archaeopteryx* consists of bilaterally symmetrical pennaceous feathers, possessing long, thin and slightly curved rachides with regularly oriented barbs, forming narrow vanes (Fig. 2d). Parts of the neck plumage

had separated from the cervical vertebrae with the skull and are preserved between the vertebral column and the skull remains (Fig. 1 and Extended Data Fig. 3); they show that pennaceous feathers also covered the neck up to the head.

Long hindlimb feathers are preserved along the femur and tibiotarsus (Fig. 2b, c). The length of these feathers is about 40–45 mm (51–58% of tibiotarsus length), but decreases from the middle to the distal end of the tibiotarsus to about 29 mm. More than 18 feathers are preserved along the right tibiotarsus; they are arranged mainly in parallel and preserved with their long axes oriented almost perpendicularly towards the tibiotarsus. The narrow vanes are fully symmetrical in shape, and seem to expand slightly towards their distal end, although this might be an artefact due to overlapping in the more basal parts. As in the body feathers, the rachides are slightly curved. In the region of the right ankle and proximal metatarsus, a row of very short (about 25% of the length of the feathers on the mid-length of the tibiotarsus), narrow-vaned, symmetrical feathers is preserved (Fig. 2b, c and Extended Data Fig. 4d), unlike the elongate pennaceous feathers in other paravians^{3,4,15}.

The disarticulated wing shows the dorsal aspect of the plumage and possesses 12 primary remiges (Fig. 2a and Extended Data Figs 4 and 5). A visible gap between the second and third primaries is artificial and caused by a disruption to the wing by the left foot during burial. The distal ends of the rachides are extremely thin, but expand abruptly approximately 20–25% of the feather length towards their proximal

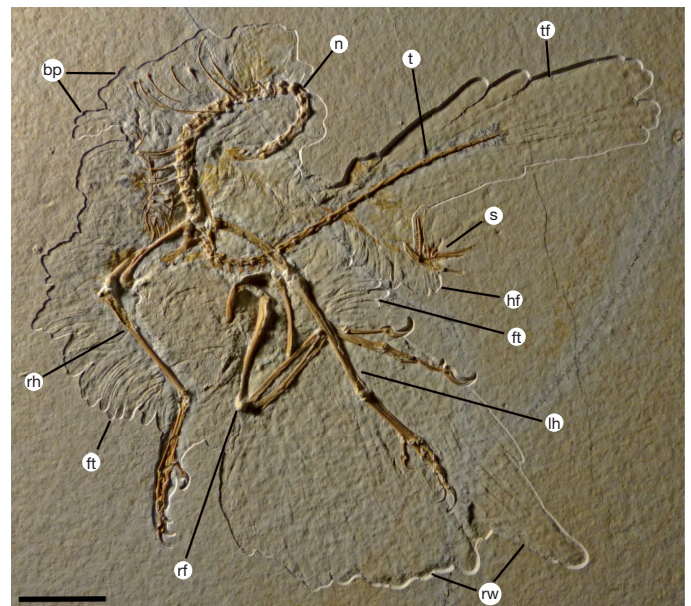


Figure 1 | Overview photograph of the 11th skeletal specimen of *Archaeopteryx*. Abbreviations: bp, body plumage; ft, feather ‘trousers’; hf, hackle feathers; lh, left hindlimb; n, neck; rf, right forelimb; rh, right hindlimb; rw, right wing; s, skull remains; t, tail; tf, tail feathers. Scale bar, 5 cm.

¹Staatliche Naturwissenschaftliche Sammlungen Bayerns, Bayerische Staatssammlung für Paläontologie und Geologie, Richard-Wagner-Straße 10, 80333 Munich, Germany. ²Department of Earth and Environmental Sciences and GeoBioCenter, Ludwig-Maximilians-Universität München, Richard-Wagner-Straße 10, 80333 Munich, Germany. ³Tannenweg 16, 85134 Stammham, Germany.

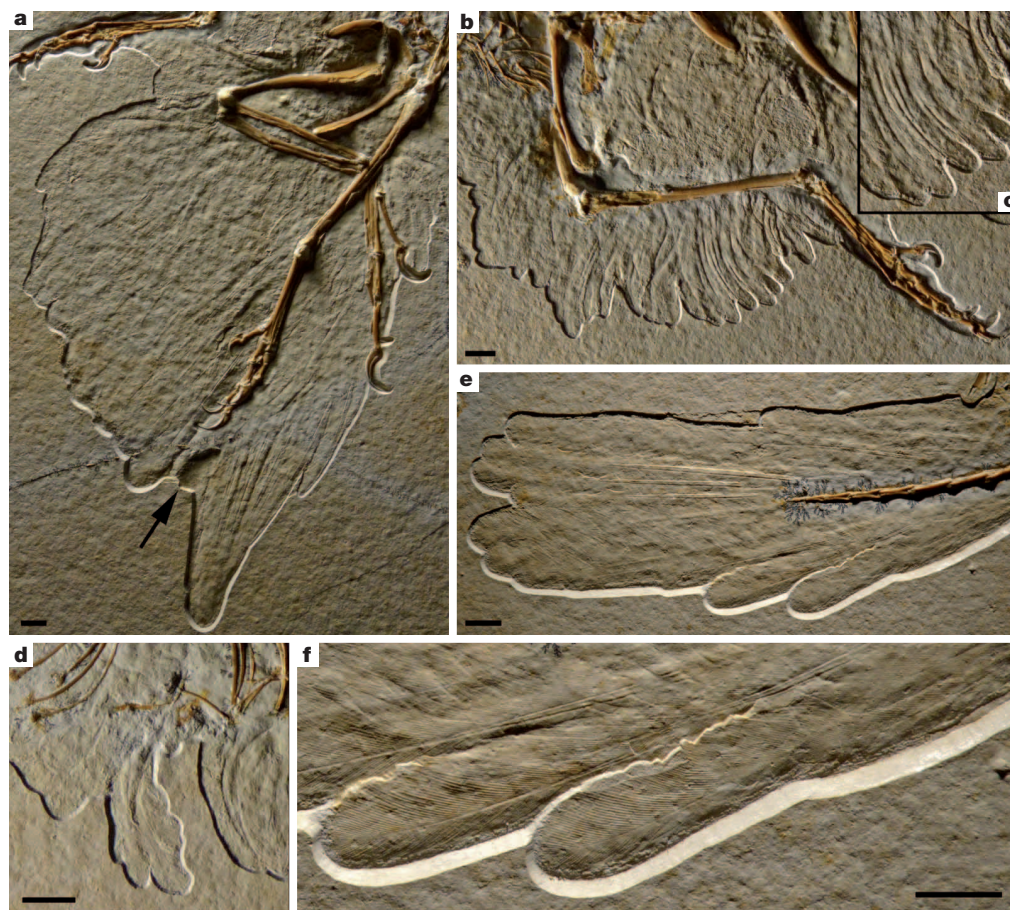


Figure 2 | Details of the plumage of the 11th specimen of *Archaeopteryx*. **a**, Right wing from dorsal view. **b**, Leg feathers of the right hindlimb. **c**, Detail of leg feathers of right hindlimb. **d**, Detail of body plumage from the belly region. **e**, Overview of tail feathers. **f**, Detail of asymmetrical feathers at the lateral side of the tail. Arrow in **a** indicates gap in the wing caused by the overlapping left foot. Scale bars, 1 cm.

parts. Dorsal coverts are preserved above the primary feathers and measure approximately half their length, as in extant birds¹⁶. The actual number of secondaries cannot be estimated owing to overlapping with coverts. These finds contradict an interpretation of the plumage of *Archaeopteryx* based on the Berlin specimen, according to which the number of primaries is reduced and the coverts are extremely elongated⁹. Thus, the wing morphology of *Archaeopteryx* conforms to that of modern birds, as previously suggested¹ (see Extended Data Fig. 5 and Supplementary Information).

The remiges of the new specimen of *Archaeopteryx* show robust rachides, indicating that recent studies assuming limited flight ability in *Archaeopteryx* and basal birds based on the relative rachis diameter⁸ might be in error owing to the poorer preservation quality of the feathers measured (see Supplementary Information and Extended Data Fig. 6).

For the first time the distal end of the tail is completely preserved in *Archaeopteryx* (Fig. 2e). The distal rectrices are symmetrical in shape and extremely long, measuring about 99–114 mm (up to 61% of the length of the bony tail). The rachis morphology resembles that of the wing feathers. The preserved shape of the tail plumage is probably narrower than originally, because some rectrices are completely covered by others. The distal end of the tail is bifurcated (Fig. 2e), but it cannot be ruled out that this might be caused by a moulting process of the terminal feather, which would indicate a centrifugal moulting pattern as in recent birds¹⁶. Lateral rectrices are slightly shorter than distal ones and clearly asymmetrical in shape (Fig. 2f). This may suggest that the tail of *Archaeopteryx* possessed an additional aerodynamic function increasing the total lift of the animal^{17,18}.

To trace the evolution of pennaceous feathers across the origin of birds, we mapped their occurrences onto a phylogeny of advanced maniraptoran theropods (Fig. 3, Extended Data Fig. 7 and Supplementary Table 2). Our analysis places *Archaeopteryx* at the base of Avialae, but

more derived than a clade formed by *Pedopenna*, *Eosinopteryx*, *Anchiornis* and *Xiaotingia*. Furthermore, we found Troodontidae as the sister group of Avialae, making Deinonychosauria (as traditionally conceived to include Dromaeosauridae and Troodontidae) paraphyletic (Extended Data Fig. 7), as recently hypothesized⁶.

Our results indicate that the distribution of pennaceous feathers is highly variable even before the advent of flight, indicating that different functional aspects were involved in the early evolution of these structures. Symmetrical pennaceous body feathers were present in the ancestor of the clade containing Oviraptorosauria and Paraves (Pennaraptora; see Supplementary Information). However, this might not be the 'true' origin of this feather type, because the morphology of body feathers in basal coelurosaurs is often uncertain owing to poor preservation or strong overlap of single feathers within the plumage¹⁹. Elongated, symmetrically shaped pennaceous remiges first appear in the stem species of Pennaraptora, whereas asymmetrical aerodynamically shaped remiges were convergently developed in the dromaeosaurid *Microraptor*³ and the common ancestor of *Archaeopteryx* and Pygostylia^{10,20}. Alula feathers allowing high manoeuvrability during low-speed flight in extant birds²¹ were convergently developed in *Microraptor* and Ornithothoraces^{3,22}.

Elongated pennaceous feathers along the tibia are found in *Microraptor*³, *Sinornithosaurus*²³ and many basal avialans^{4,5,15}, indicating their presence in the last common ancestor of Dromaeosauridae, Troodontidae and Avialae. In contrast, elongated pennaceous feathers on the metatarsus are only known for *Microraptor*³, *Xiaotingia*⁵, *Anchiornis*⁴, *Pedopenna*²⁴ and *Sapeornis*¹⁵. Based on our tree topology this feature did not originate singularly within Paraves, but evolved three times independently, in Dromaeosauridae, *Sapeornis* and the clade containing *Anchiornis*, *Pedopenna* and *Xiaotingia*.

Pennaceous rectrices were also present in the common ancestor of Pennaraptora, but the original shape of the tail plumage and its evolution

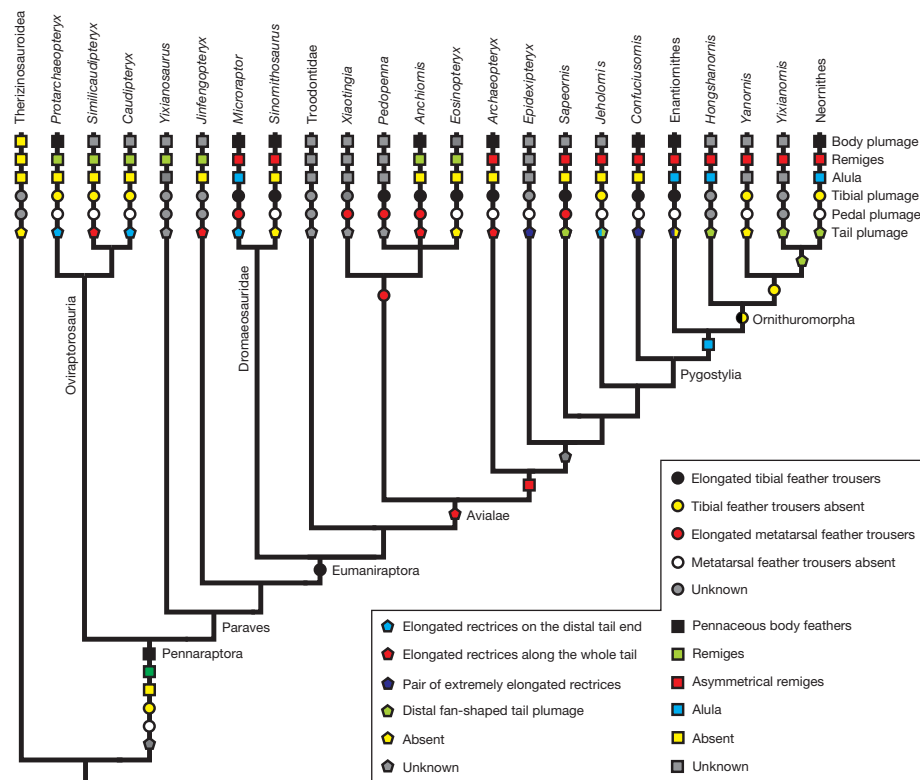


Figure 3 | Evolution of pennaceous feathers in maniraptoran theropods. Several character states referring to the presence/absence of pennaceous feathers on different parts of the body mapped onto a phylogenetic hypothesis

of maniraptoran theropods. For details of phylogenetic analysis see Supplementary Information.

are ambiguous. In *Protarchaeopteryx*, *Caudipteryx* and *Microraptor*, elongated pennaceous rectrices are present in the distal half of the tail^{2,12}, whereas *Jeholornis* seems to be unique in showing a proximal and a distal tail fan²⁵. The tail of *Similicaudipteryx*, *Jinfengopteryx*, *Anchiornis* and *Archaeopteryx* is completely covered with such feathers^{4,20,26}, whereas *Eosinopteryx* lacks them completely²⁷. Within short-tailed avialans, elongated rectrices are absent in some Enantiornithes, *Yanornis* and many *Confuciusornis* specimens^{15,28}. In contrast, tail plumages with a pair of extremely elongated pennaceous feathers are known for *Confuciusornis*, some Enantiornithes and *Epidexipteryx*^{28,29}. A fan-shaped tail plumage as present in modern birds is convergently developed in *Sapeornis* and several Ornithuromorpha^{15,30}.

Assuming a single origin of pennaceous feathers within theropods, the present data indicate that this feather type evolved in the common ancestor of Pennaraptora, but rapidly differentiated into various subtypes within different body regions, leading to a considerable plumage diversity compared with more basal theropods. Based on these results, the evolution of pennaceous feathers is generally decoupled from the origin of flight and might be related to other biological roles. Like the filamentous plumages of more basal theropods, pennaceous body feathers were certainly important for body insulation. However, these feathers may have further been used for camouflage and display^{7,11}. Furthermore, elongated remiges on the forelimb were probably used for balance during locomotion and breeding¹⁰. Although some taxa possess prominent feather trousers, an initial aerodynamic function, as previously hypothesized^{3,13}, can be rejected because the presence of this character is extremely variable within Paraves, and elongated hindlimb feathers were usually restricted to the tibia and are symmetrical in shape. Thus, probably non-volant taxa such as *Anchiornis* or *Xiaotingia* might have used these feathers for display, breeding or other functions. The display function might be supported by the finding of complex colour or iridescent patterns in hindlimb feathers^{11,12}. *Archaeopteryx* might have used its 'trouser' additionally as a vertical plane during landing, similar to recent raptors. Only for *Microraptor* does an improved

aerodynamic adaptation of the trousers seem plausible, as indicated by the triangular shape of the trousers and the asymmetrical shape of the feathers. These results contradict the hypothesis that the flapping flight of modern birds was preceded by a four-winged gliding stage^{3,13}, and indicates that flight ability in Avialae and *Microraptor* evolved convergently and was functionally different.

The huge diversity and homoplasy of plumage morphology of the tail within Pennaraptora indicate that this character may initially have evolved for display^{10,12,29}. The asymmetrical shape of the lateral rectrices in *Archaeopteryx* and ornithuromorphs indicates a secondary aerodynamic function for the tail.

The current results show that pennaceous feathers were present in the common ancestor of Pennaraptora. As in extant birds, this feather type probably fulfilled several biological roles related to body insulation, manoeuvrability, brooding, camouflage and display^{7,16}. However, given the great diversity of pennaceous feathers found within different body regions and across the phylogeny, it seems plausible that the evolution of this feather type (especially in the wing, hindlimbs and tail) was primarily driven by display functions. These structures then represented an exaptation that was repeatedly, and probably convergently, recruited for aerodynamic functions. This indicates that the origin and evolution of flight in theropod dinosaurs were more complex than previously thought, and (as already shown for other anatomical complexes, for example forelimb myology and breathing apparatus) could draw on structures that evolved in different functional contexts.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 January; accepted 12 May 2014.

- Wellnhofer, P. *Archaeopteryx. Der Urvogel von Solnhofen* (Dr Friedrich Pfeil, 2008).
- Ji, Q., Currie, P. J., Norell, M. A. & Ji, S. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
- Xu, X. *et al.* Four-winged dinosaurs from China. *Nature* **421**, 335–340 (2003).

4. Hu, D., Hou, L., Zhang, L. & Xu, X. A pre-*Archaeopteryx* troodontid theropod from China with long feathers on the metatarsus. *Nature* **461**, 640–643 (2009).
5. Xu, X., You, H., Du, K. & Han, F. An *Archaeopteryx*-like theropod from China and the origin of Avialae. *Nature* **475**, 465–470 (2011).
6. Godefroit, P. *et al.* A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. *Nature* **498**, 359–362 (2013).
7. Prum, R. O. & Brush, A. H. The evolutionary origin and diversification of feathers. *Q. Rev. Biol.* **77**, 261–295 (2002).
8. Nudds, R. L. & Dyke, G. J. Narrow primary feathers rachises in *Confuciusornis* and *Archaeopteryx* suggest poor flight ability. *Science* **328**, 887–889 (2010).
9. Longrich, N. R., Vinther, J., Meng, Q., Li, Q. & Russell, A. P. Primitive wing feather arrangement in *Archaeopteryx lithographica* and *Anchiornis huxleyi*. *Curr. Biol.* **22**, 1–6 (2012).
10. Xu, X. & Guo, Y. The origin and early evolution of feathers: insights from recent paleontological and neontological data. *Vertebrata Palasiatica* **47**, 311–329 (2009).
11. Li, Q. *et al.* Plumage color patterns of an extinct dinosaur. *Science* **327**, 1369–1372 (2010).
12. Li, Q. *et al.* Reconstruction of *Microraptor* and the evolution of iridescent plumage. *Science* **335**, 1215–1219 (2012).
13. Longrich, N. R. Structure and function of hindlimb feathers in *Archaeopteryx lithographica*. *Paleobiology* **32**, 417–431 (2006).
14. Turner, A. H., Makovicky, P. J. & Norell, M. A. A review of dromaeosaurid systematics and paravian phylogeny. *Bull. Am. Mus. Nat. Hist.* **371**, 1–206 (2012).
15. Zheng, X. *et al.* Hind wings in basal birds and the evolution of leg feathers. *Science* **339**, 1309–1312 (2013).
16. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy. Integument, Parts I and II* 197–233 (U.S. Government Printing Office, 1972).
17. Gatesy, S. M. & Dial, K. P. From frond to fan: *Archaeopteryx* and the evolution of short-tailed birds. *Evolution* **50**, 2037–2048 (1996).
18. Meseguer, J. *et al.* Lift devices in the flight of *Archaeopteryx*. *Span. J. Palaeontol.* **27**, 125–130 (2012).
19. Foth, C. On the identification of feather structures in stem-line representatives of birds: evidence from fossils and actuopalaeontology. *Paläontol. Z.* **86**, 91–102 (2012).
20. Xu, X., Zheng, X. & You, H. Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010).
21. Nachtigall, W. & Kempf, B. Vergleichende Untersuchungen zur flugbiologischen Funktion des Daumenfittichs (*Alula spuria*) bei Vögeln. *Z. Vgl. Physiol.* **71**, 326–341 (1971).
22. Chiappe, L. M. & Dyke, G. J. The Mesozoic radiation of birds. *Annu. Rev. Ecol. Syst.* **33**, 91–124 (2002).
23. Ji, Q., Norell, M. A., Gao, K., Ji, S. & Ren, D. The distribution of integumentary structures in a feathered dinosaur. *Nature* **410**, 1084–1088 (2001).
24. Xu, X. & Zhang, F. A new maniraptoran dinosaur from China with long feathers on the metatarsus. *Naturwissenschaften* **92**, 173–177 (2005).
25. O'Connor, J. *et al.* Unique caudal plumage of *Jeholornis* and complex tail evolution in early birds. *Proc. Natl Acad. Sci. USA* **110**, 17404–17408 (2013).
26. Ji, Q. *et al.* First avialan bird from China. *Geol. Bull. China* **24**, 197–210 (2005).
27. Godefroit, P. *et al.* Reduced plumage and flight ability of a new Jurassic paravian theropod from China. *Nature Commun.* **4**, 1–6 (2013).
28. Chiappe, L. M., Marugán-Lobón, J., Ji, S. & Zhou, Z. Life history of a basal bird: morphometrics of the Early Cretaceous *Confuciusornis*. *Biol. Lett.* **4**, 719–723 (2008).
29. O'Connor, J. K., Chiappe, L. M., Chuong, C., Bottjer, D. J. & You, H. Homology and potential cellular and molecular mechanisms for the development of unique feather morphologies in early birds. *Geosciences* **2**, 157–177 (2012).
30. Chiappe, L. M. *et al.* A new specimen of the Early Cretaceous bird *Hongshanornis longicresta*: insights into the aerodynamics and diet of a basal ornithuromorph. *PeerJ* **2**, e234 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Pohl for making the specimen available for study, C. Keilmann and R. Albersdörfer for assistance, Xu X. for access to material, and R. Carney, G. Mayr and A. López-Arbarelo for discussions. This study was supported by the Volkswagen Foundation under grant I/84 640 (to O.W.M.R.).

Author Contributions C.F. and O.W.M.R. designed the study, collected and analysed the data and wrote the paper; H.T. did all the photography and helped with discussions.

Author Information The data matrix for the phylogenetic analysis has been deposited in MorphoBank (<http://www.morphobank.org>) as project 1126. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.W.M.R. (o.rauhut@lrz.uni-muenchen.de).

Genomic divergence in a ring species complex

Miguel Alcaide^{1,2}, Elizabeth S. C. Scordato^{3,4}, Trevor D. Price³ & Darren E. Irwin¹

Ring species provide particularly clear demonstrations of how one species can gradually evolve into two, but are rare in nature^{1–3}. In the greenish warbler (*Phylloscopus trochiloides*) species complex, a ring of populations wraps around Tibet. Two reproductively isolated forms co-exist in central Siberia, with a gradient of genetic and phenotypic characteristics through the southern chain of populations connecting them^{4–6}. Previous genetic evidence has proven inconclusive, however, regarding whether species divergence took place in the face of continuous gene flow and whether hybridization between the terminal forms of the ring ever occurred^{7–9}. Here we use genome-wide analyses to show that, although spatial patterns of genetic variation are currently mostly as expected of a ring species, historical breaks in gene flow have existed at more than one location around the ring, and the two Siberian forms have occasionally interbred. Substantial periods of geographical isolation occurred not only in the north but also in the western Himalayas, where there is now an extensive hybrid zone between genetically divergent forms. Limited asymmetric introgression has occurred directly between the Siberian forms, although it has not caused a blending of those forms, suggesting selection against introgressed genes in the novel genetic background. Levels of reproductive isolation and genetic introgression are consistent with levels of phenotypic divergence around the ring, with phenotypic similarity and extensive interbreeding across the southwestern contact zone and strong phenotypic divergence and nearly complete reproductive isolation across the northern contact zone. These results cast doubt on the hypothesis that the greenish warbler should be viewed as a rare example of speciation by distance⁶, but demonstrate that the greenish warbler displays a continuum from slightly divergent neighbouring populations to almost fully reproductively isolated species.

Ring species are defined as a chain of populations wrapping around a geographical barrier, with only one location where two reproductively isolated forms co-occur. In the ideal case, the two terminal forms do not exchange genes directly, but are connected by a chain of interbreeding populations^{1,10}. Despite their rarity^{2–4}, ring species have been integral to recent studies of speciation, first because continuous spatial variation can be used as a means to understand how divergence proceeds in time¹, and second because they may provide examples of speciation despite gene flow^{4,10,11}. The frequency with which progress to speciation takes place in the face of gene exchange is uncertain^{8,11,12}. It has been suggested¹³ with respect to ring species that their formation may have involved periods of allopatry, during which gene flow was interrupted, interspersed with periods of contact and gene exchange. This seems especially likely to be involved in recent avian speciation events, given the cyclic climatic upheavals during the Pleistocene and the time frame, around two million years, over which parapatric bird taxa continue to form hybrid zones in nature^{11,14}.

The greenish warbler *P. trochiloides* encompasses six different recognized subspecies, of which five (*P. t. viridanus* in west Siberia, *P. t. ludlowi* in the western Himalayas, *P. t. trochiloides* in the central and eastern Himalayas, *P. t. obscuratus* in China and *P. t. plumbeitarsus* in eastern Siberia) form a ring around the inhospitable Tibetan plateau, and the sixth, *P. t. nitidus*, is isolated in the Caucasus (Fig. 1, Extended Data Fig. 1).

A gradient in morphology through this ring of populations has been described¹⁵, with the only sharp break being between *viridanus* and *plumbeitarsus* to the north of Tibet. Further study has shown that song, calls and plumage differ discretely between these two Siberian forms and interbreeding has never been observed^{4,5}. However, some mixed singers^{9,16} and anecdotal observation of territorial interactions between a *viridanus* and its *plumbeitarsus* neighbours (D.E.I., unpublished observation) suggest the possibility of some hybridization.

Genetic evidence has proved inconclusive regarding whether breaks in gene flow occurred anywhere other than between the northern forms^{4,6}. On the one hand, mitochondrial DNA sequences indicate several possible phylogeographic breaks around the ring⁴. Most notably, two major mtDNA clades separating the terminal forms in Siberia overlap over roughly 100 km in the western Himalayas (see Fig. 1), in the region designated¹⁵ as the range of *ludlowi*. On the other hand, nuclear DNA as assessed using AFLP (amplified fragment length polymorphism) markers shows a smooth gradient of genetic characteristics around the ring⁶ and, despite an apparent steepening of genetic variation across the southwestern mitochondrial split, the AFLP patterns fit an isolation-by-distance model well^{17,18}. Such discrepancies may be reconciled in two ways. First, stronger phylogeographic structure may arise in the uniparentally inherited mtDNA even with no geographical barriers because of its lower effective population size, non-recombining inheritance and likelihood of being exposed to geographically varying natural selection¹⁹. Second, genetic structure may be difficult to detect in AFLP markers owing to limited marker resolution coupled with the possibility of gene flow following secondary contact. In this study, we take advantage of the increased genomic resolution provided by a next-generation sequencing (NGS) approach²⁰ to (1) assess evidence for ancient periods of allopatric divergence by asking if phylogeographic breaks in mtDNA correspond to sharp transitions in nuclear DNA and (2) assess levels of recent genetic exchange across these historical breaks.

We genotyped 95 greenish warblers sampled from 22 sites (Extended Data Fig. 1 and Extended Data Table 1) at 2,334 single nucleotide polymorphisms (SNPs; Supplementary Information file 1). Bayesian analysis identifies five ancestral genetic clusters (Fig. 1a, Extended Data Fig. 1); one of these corresponds to allopatric *nitidus* from the Caucasus, leaving four around the ring. The *viridanus* and *trochiloides* clusters intergrade along the range of *ludlowi*. This gradient supports the assertion¹⁵ that *ludlowi* is a transitional form; both western and eastern mitochondrial types co-occur in the centre of the transition zone (Fig. 1a) and this subspecies also exhibits the lowest proportion of private alleles of any taxon in the ring (Extended Data Table 2). A deeper examination of the southwestern transition zone using 4,018 SNPs and an additional 39 birds sampled from its centre (Supplementary Information file 2) reveals the features of a classic hybrid zone²¹ with a cline width estimated at 220 km (Fig. 1b). This cline is wider than those between old hybridizing species, where selection against hybrids is presumed strong^{11,21}. The east side of the ring has been sampled at fewer sites than the west. Here, the *obscuratus* subspecies appears as an intergrading form between *trochiloides* and *plumbeitarsus*, but it also displays a considerable ancestry of its own (Fig. 1a).

On the north side of the ring, the Bayesian analysis shows *viridanus* and *plumbeitarsus* as distinct genomic clusters. Unexpectedly, the analysis also

¹Department of Zoology and Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. ²Department of Evolutionary Ecology, Estación Biológica de Doñana – CSIC, Americo Vespucio s/n, 41092 Sevilla, Spain. ³Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. ⁴Department of Ecology and Evolutionary Biology, The University of Colorado, Boulder, Colorado 80309, USA.

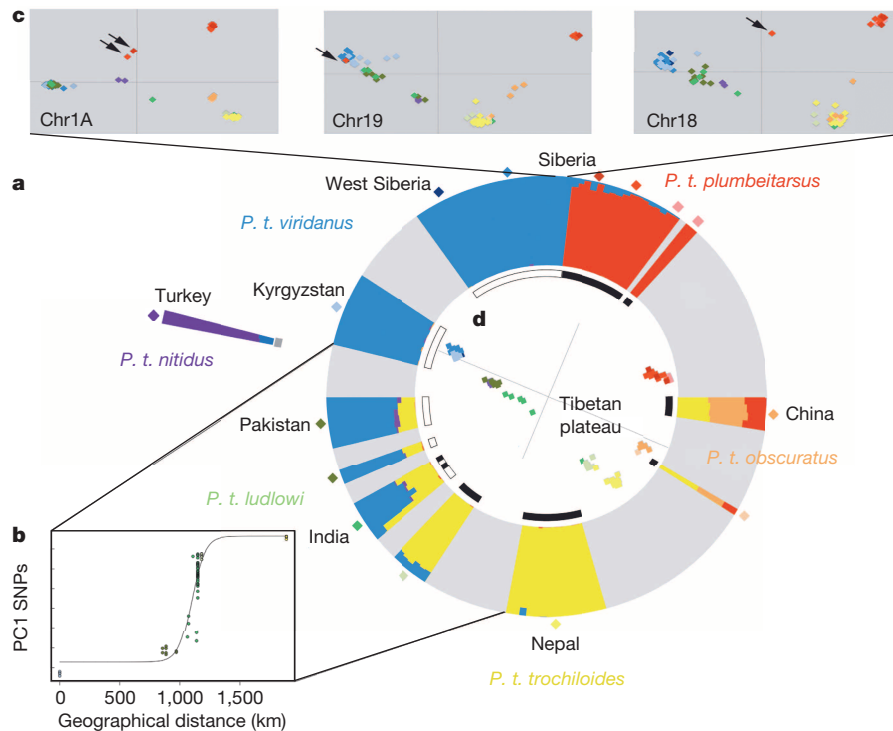


Figure 1 | Spatial patterns of genetic differentiation within the greenish warbler ring species. **a**, Bayesian clustering analysis (BCA) of 95 greenish warblers illustrated as a ring corresponding roughly to geography. For each individual, coloured bars indicate the ancestry proportions for each of the five inferred genetic clusters (see also Extended Data Fig. 4). Light grey areas within the ring depict gaps in sampling (see Extended Data Fig. 1 for details). Inner bars indicate mtDNA types according to ref. 4: white bars, western clade; black bars, eastern clade. Coloured diamonds around the ring designate sampling sites. **b**, Spatial distribution of nuclear genetic variation along the range of *P. t. ludlowi* ($N = 66$ birds) suggests an extensive hybrid zone between *P. t. viridanus* and *P. t. trochiloides*. Colours correspond to the sampling sites (coloured diamonds) surrounding the BCA plot. **c**, Geographical variation

indicates that up to 10% of the genome of some *plumbeitarsus* individuals consists of material introgressed from *viridanus*; such asymmetry in introgression has been commonly observed in avian hybrid zones²². The sampling location furthest from the contact zone is the only region where *plumbeitarsus* show no signal of introgression (Fig. 1a, Extended Data Fig. 1). Principal coordinate analysis (PCA) using sets of highly differentiated SNPs (fixation index $F_{ST} \geq 0.85$, a measure of population differentiation) between western and eastern Siberian populations, and grouped by chromosome, illustrates the introgression of large blocks of DNA. Thus, in a small number of chromosomes per individual, some *plumbeitarsus* birds exhibit intermediate genetic characteristics between *viridanus* and *plumbeitarsus* (that is, heterozygotes for most diagnostic markers that distinguish the two taxa) or even cluster with *viridanus* (that is, homozygotes for the foreign chromosome type; Fig. 1c, see also Supplementary Information file 3). This pattern is best explained by recent introgression of large blocks of DNA, because there is little evidence for recombination resulting in the breakdown of these blocks.

Nevertheless, genome-wide PCA using 2,334 SNPs (Fig. 1d) depicts the two northern forms as clearly distinct genetic clusters, both connected around the southern side of the ring by a gradient of genetic characteristics. Similarly to AFLP studies⁶, genetic divergence and geographical distance around the ring are strongly correlated (Mantel's $r = 0.71$, $P < 0.0001$, see Extended Data Figs 2 and 3). Unlike the AFLP analysis, however, the much greater resolution of the present genomic analysis reveals significant discontinuities in the gradient of genetic variation. Genomic breaks between Pakistan and Kyrgyzstan, Nepal and China, or China and eastern Siberia may be the result of geographical gaps in

and evidence for introgression in a set of highly differentiated SNP markers ($F_{ST} \geq 0.85$) from three different chromosomes (1A, 18 and 19) between western and eastern Siberian greenish warblers, as summarized by principal coordinate analyses (PCA, $N = 95$ birds). Arrows pinpoint *plumbeitarsus* individuals clustering with *viridanus* or showing intermediate genetic characteristics between the main *viridanus* and *plumbeitarsus* clusters. **d**, Genome-wide PCA based on the analysis of 2,334 SNPs in 95 greenish warblers. PC1 explains 58.45% of the variance and PC2 22.07%. *P. t. viridanus* and *P. t. plumbeitarsus* are strongly differentiated by both PCs, with other populations genetically intermediate. The two *nitidus* individuals strongly differentiate from the other western individuals across the third dimension of the PCA (not shown here).

sampling and could be consistent with either isolation by distance or historical barriers; more detailed geographical sampling would be required to discriminate between these possibilities. On the contrary, the abrupt transition of genetic characteristics in the southwest cannot be attributed to limited sampling and is best explained by secondary contact and hybridization between groups that diverged in allopatry (Fig. 1b).

Our results indicate that allopatric divergence played a strong role in shaping patterns of genomic divergence during the formation of the greenish warbler ring. Previous findings nonetheless support the idea that geographical isolation has not been the sole driver for the establishment of reproductive isolation; natural and sexual selection also appear to play major roles^{4,5,23}. Divergence patterns of traits under selection in greenish warblers are in accord with much stronger reproductive isolation across the northern break than across the southern. Song and call divergence between *viridanus* and *plumbeitarsus*, for example, has contributed to pre-mating isolation to the extent that the two taxa discriminate between each other's songs^{4,5}. Songs on either side of the southwestern contact zone, by contrast, are very similar^{4,16}. Considerations of biogeography and the genetics of migration in other species also suggest that hybrids in the north might have an intermediate migratory route, causing them to attempt to migrate over Tibet rather than along more optimal routes around it²³. If any migratory divide exists in the south it is less likely to drive a reduction in hybrid fitness, given that intermediate migration routes would take hybrids into India, rather than Tibet.

In summary, genomic relationships among populations of greenish warblers are mostly as expected of a ring species, with the terminal forms in Siberia being markedly different, with the intervening populations

connected by a genetic gradient. However, the present analysis shows, first, that there have been periods of population separation in at least one location other than the meeting point of the terminal forms, and, second, that genetic exchange has occurred at each area of secondary contact. Introgression across the northern break indicates that barriers to interbreeding are not absolute. The divergence time of the two principal mtDNA clades in the greenish warbler complex dates to 1–2 Myr ago¹¹, a span of time during which hybrid fertility is common between diverging bird populations²⁴. Introgression is in accord with the observation that ~10% of all bird species, many as old or older than greenish warblers^{11,14}, have been recorded as entering into hybrid pairings^{25,26}; several have now been shown to have portions of their genome introgressed²². Rare hybridization between terminal forms of a ring species should be expected, given what we know about hybrid zone dynamics and the timescale of species formation¹¹. The two Siberian forms, however, are unlikely to blend together despite gene flow.

Finally, this study provides meaningful information concerning the interpretation of ring species as prominent illustrations of speciation in the face of gene flow^{1–8,10,11,27}. Earlier analyses of AFLP markers supported the greenish warbler as the last known example in birds of an ideal ring species in which the terminal forms became reproductively isolated despite being connected by gene flow during their whole history of divergence^{1,6,17,18}. It is possible that speciation-by-distance ring species could exist, but their extreme rarity may be explained by the rapidity of Earth's climatic shifts compared to the time that reproductive isolation takes to occur; over these great spans of time, and as exemplified by the present study, populations are very likely to be temporarily divided. Nonetheless, the greenish warbler and other 'broken' ring species^{1–6} show how gradual genetic and phenotypic changes can build up to produce the differences between distinct species in the way that they were first envisioned^{10,13,27}. The quest must continue for any exemplar that formed with continuous gene flow throughout its entire history.

METHODS SUMMARY

Next-generation sequencing. We used genomic DNA from adult greenish warblers captured during the course of the studies of refs 4 and 6. In addition, a second library was constructed from genomic DNA of a different sample of 39 adult birds from Keylong, Himachal Pradesh (India). Library preparations were executed following a genotyping-by-sequencing²⁰ (GBS) protocol and were run into independent lanes of the Illumina HiSeq 2000 automated sequencer according to the paired-end sequencing protocol.

De novo discovery and genotyping of SNPs. We implemented an analysis workflow that was assisted by the zebra finch *Taeniopygia guttata* genome²⁸ and an incomplete genome draft of the greenish warbler. Out of 81,034 candidate SNPs, we chose those for which a minimum of three reads per SNP marker were retrieved in all individuals, with an average coverage per site larger than 16 reads (Extended Data Table 3), resulting in data sets of 2,334 SNPs for the first set of samples (around the whole ring) and 4,018 for the second (across the southwestern contact zone).

Genetic structuring. Principal coordinate plots and basic estimates of population differentiation (F_{ST}) and genetic diversity were generated in the GenAlEx 6.5 package²⁹. Genetic admixture was investigated through the Bayesian model-based clustering method implemented in STRUCTURE version 2.3.4³⁰.

Hybrid zone analysis. The width of genomic clines was estimated by fitting a sigmoid curve to the relationship between geographical location and the primary axis of genomic variation, according to the sigmoid equation provided by ref. 21, in R.

Introgression analysis. We selected a subset of markers showing high genetic differentiation ($F_{ST} \geq 0.85$) between *viridanus* and *plumbeitarsus*, grouped 1,827 SNPs by chromosome, and then used PCA to inspect patterns of genetic relationships for each chromosome.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January; accepted 26 March 2014.

Published online 25 May 2014.

1. Irwin, D. E., Irwin, J. H. & Price, T. D. Ring species as bridges between microevolution and speciation. *Genetica* **112–113**, 223–243 (2001).

2. Moritz, C., Schneider, C. J. & Wake, D. B. Evolutionary relationships within the *Ensantina eschscholtzii* complex confirm the ring species interpretation. *Syst. Biol.* **41**, 273–291 (1992).
3. Cacho, N. I. & Baum, D. A. The Caribbean slipper spurge *Euphorbia tithymaloides*: the first example of a ring species in plants. *Proc. R. Soc. B* **279**, 3377–3383 (2012).
4. Irwin, D. E., Bensch, S. & Price, T. D. Speciation in a ring. *Nature* **409**, 333–337 (2001).
5. Irwin, D. E., Thimman, M. P. & Irwin, J. H. Call divergence is correlated with geographic and genetic distance in greenish warblers (*Phylloscopus trochiloides*): a strong role for stochasticity in signal evolution? *J. Evol. Biol.* **21**, 435–448 (2008).
6. Irwin, D. E., Bensch, S., Irwin, J. H. & Price, T. D. Speciation by distance in a ring species. *Science* **307**, 414–416 (2005).
7. Wake, D. B. Speciation in the round. *Nature* **409**, 299–300 (2001).
8. Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, 2004).
9. Kovylov, N. S., Marova, I. M. & Ivanitskii, V. V. Variation of song and plumage in the western (*Phylloscopus trochiloides viridanus*) and eastern (*Phylloscopus trochiloides plumbeitarsus*) forms of the greenish warbler in a sympatry zone: is the hypothesis of ring speciation true? *Biol. Bull.* **39**, 729–740 (2012).
10. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, 1942).
11. Price, T. D. *Speciation in Birds* (Roberts and Co., 2008).
12. Endler, J. A. *Geographic Variation, Speciation, and Clines* (Princeton Univ. Press, 1977).
13. Mayr, E. *Animal Species and Evolution* (Belknap, 1963).
14. Weir, J. T. & Price, T. D. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *Am. Nat.* **177**, 462–469 (2011).
15. Ticehurst, C. B. *A Systematic Review of the Genus Phylloscopus* (Trustees of the British Museum, London, 1938).
16. Irwin, D. E. in *Creating Consilience: Integrating the Sciences and the Humanities* (eds Slingerland, E. & Collard, M.) 163–178 (Oxford Univ. Press, 2012).
17. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nature Genet.* **40**, 646–649 (2008).
18. Martins, A. B., Aguiar, M. A. M. & Bar-Yam, Y. Evolution and stability of ring species. *Proc. Natl Acad. Sci. USA* **110**, 5080–5084 (2013).
19. Irwin, D. E. Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *Am. Nat.* **180**, 35–49 (2012).
20. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
21. Barton, N. H. & Hewitt, G. M. Adaptation, speciation and hybrid zones. *Nature* **341**, 497–503 (1989).
22. Rheindt, F. E. & Edwards, S. V. Genetic introgression: an integral but neglected component of speciation in birds. *Auk* **128**, 620–632 (2011).
23. Irwin, D. E. & Irwin, J. H. in *Birds of Two Worlds: The Ecology and Evolution of Migration* (eds Greenberg, R. & Marra, P. P.) 27–40 (John Hopkins Univ. Press, 2005).
24. Price, T. D. & Bouvier, M. M. The evolution of F_1 postzygotic incompatibilities in birds. *Evolution* **56**, 2083–2089 (2002).
25. McCarthy, E. M. *Handbook of Avian Hybrids of the World* (Oxford Univ. Press, 2006).
26. Grant, P. R. & Grant, B. R. Hybridization of bird species. *Science* **256**, 193–197 (1992).
27. Dobzhansky, T. in *A Century of Darwin* (ed. Barnett, S. A.) 19–55 (Heinemann, 1958).
28. Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757–762 (2010).
29. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. *Bioinformatics* **28**, 2537–2539 (2012).
30. Pritchard, J. K., Stephens, J. K. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (project no. 273773) and an NSERC discovery grant (no. 311931). Original collections were made with a grant from the US NSF. We thank Z. Benowitz-Fredericks, J. Gibson, S. Gross, J. Irwin, G. Kelberg, A. Knorre, K. Marchetti and B. Sheldon for assistance in the field. For additional samples we thank P. Alström, K. Marchetti, U. Olsson, A. Richman and J. Tiainen. We thank J. Irwin and L. Rieseberg for comments and discussion and A. Kuzmin for assistance during library preparation.

Author Contributions M.A. and D.E.I. conceived the genomic study. M.A. designed the genomic analysis, and collected and analysed the sequence data. D.E.I., E.S.C.S. and T.D.P. conducted fieldwork. D.E.I. conducted the cline analysis in the southwestern contact zone. All authors contributed to the writing of the manuscript and have read and approved the manuscript.

Author Information Raw Illumina sequencing data have been deposited in the Sequence Read Archive (SRA) under accession numbers SRX472921 and SRX473141. The list of contigs and information concerning the physical location of the SNP markers used in this study, using the zebra finch genome assembly as reference, have been deposited in the Dryad repository (<http://doi.org/10.5061/dryad.6kn93>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.I. (irwin@zoology.ubc.ca).

RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast

JongDae Shin¹, Mary C. Wallingford², Judith Gallant³, Chelsea Marcho², Baowei Jiao^{1†}, Meg Byron³, Michael Bossenz⁴, Jeanne B. Lawrence³, Stephen N. Jones³, Jesse Mager² & Ingolf Bach^{1,5}

In female mice, two forms of X-chromosome inactivation (XCI) ensure the selective silencing of female sex chromosomes during mouse embryogenesis. Beginning at the four-cell stage, imprinted XCI (iXCI) exclusively silences the paternal X chromosome. Later, around implantation, epiblast cells of the inner cell mass that give rise to the embryo reactivate the paternal X chromosome and undergo a random form of XCI (rXCI)^{1,2}. *Xist*, a long non-coding RNA crucial for both forms of XCI, is activated by the ubiquitin ligase RLIM (also known as Rnf12)^{3–5}. Although RLIM is required for triggering iXCI in mice, its importance for rXCI has been controversial. Here we show that RLIM levels are downregulated in embryonic cells undergoing rXCI. Using mouse genetics we demonstrate that female cells lacking RLIM from pre-implantation stages onwards show hallmarks of XCI, including *Xist* clouds and H3K27me3 foci, and have full embryogenic potential. These results provide evidence that RLIM is dispensable for rXCI, indicating that in mice an RLIM-independent mechanism activates *Xist* in the embryo proper.

The X-linked gene *Rlim* encodes the RING-finger ubiquitin ligase RLIM^{6,7}, which functions as a sex-specific epigenetic regulator of nurturing tissues in female mice. In the mammary glands of pregnant and lactating females, RLIM expressed from the paternal X chromosome (Xp) serves as a survival factor for milk-producing alveolar cells^{8,9}. In contrast, female embryos with a maternal deletion of *Rlim* in the germ line (Δ_m) show defective *Xist* cloud formation and subsequent Xp silencing during iXCI, and die between embryonic day (E)5.5 and E10.5 due to a lack of extraembryonic trophoblast tissues in the placenta⁵. Using an embryonic stem (ES)-cell model, evidence has been provided that RLIM can serve as a dose-dependent activator of XCI: the overexpression of RLIM induces ectopic *Xist* clouds in male and female ES cells⁴, and, in this system, homozygous disruption of *Rlim* results in a failure to initiate *Xist* transcription during rXCI^{3,10}. Thus, current models of rXCI present RLIM as a crucial activator of *Xist*^{11,12}, although some evidence suggests that it has a less important role in mice⁵.

During mouse development, RLIM-encoding messenger RNA is ubiquitously expressed, whereas its protein expression profile seems more restricted^{6,13}. Thus, to investigate *in vivo* functions of RLIM for rXCI, we first analysed its protein expression and co-stained early embryos with antibodies against RLIM and the pluripotency factor Oct4, which identifies undifferentiated cells in the inner cell mass (ICM) undergoing rXCI. Whereas RLIM seemed to be uniformly expressed in the cells of E3.5 embryos (Fig. 1a), at E4.5 and E5.5 only low levels were detected in Oct4-positive cells, in contrast to extraembryonic cell types (Fig. 1b, c and Extended Data Fig. 1a; data not shown). Likewise, at E6.5 and E7–7.5 we detected low levels of RLIM in embryonic epiblast tissues and amnion tissues, whereas levels were high in many extraembryonic tissues, including cells of the ectoplacental cone, extraembryonic ectoderm and visceral endoderm (Fig. 1d and Extended Data Fig. 1b; data not shown). Indeed, owing to low levels of RLIM, it was difficult to distinguish wild-type

embryos from those lacking RLIM in embryonic tissues by immunostaining (Extended Data Fig. 1b). Although we did not genotype E5.5 embryos for technical reasons, in all ten wild-type embryos analysed we detected no/very low levels of RLIM in Oct4-positive nuclei. Moreover, we observed similarly low RLIM levels in genotyped males and female embryos at E4.5 and E7–7.5 of various genetic backgrounds (Fig. 1 and Extended Data Fig. 1; data not shown). As rXCI occurs around E5–E5.5, our data indicate that levels of RLIM are downregulated in mouse epiblast cells before rXCI.

To induce the conditional knockout of *Rlim* (cKO) in female embryos after the occurrence of iXCI but before induction of rXCI, we used *Sox2-Cre* (SC) transgenic mice^{14,15}, as *Sox2* is specifically expressed in embryonic epiblast cells of E3.5 blastocysts¹⁶, and a paternally transmitted SC transgene is robustly transcribed in ICM cells of pre-implantation blastocysts¹⁴. Indeed, from a *fl/fl* \times WT/Y-SC (where *fl* indicates floxed, WT indicates wild type, and Y-SC indicates a male that carries both a Y chromosome and the SC transgene) cross, female and male pups with a maternal cKO (cKO_m) are born in Mendelian sex ratios, and the percentages of male and female pups carrying cKO_m were similar (Fig. 2a), confirming normal iXCI. Importantly, female pups carrying a cKO_m and a paternal germline deletion of *Rlim* (cKO_m/ Δ_p) generated via a *fl/fl* \times cKO/Y-SC cross are born at Mendelian ratios (Fig. 2b), and, except for a mammary phenotype, adult cKO_m/ Δ_p females appear normal and are fertile (data not shown). In genotyping these animals using polymerase chain reaction (PCR) we detected only the knockout allele but no longer the floxed allele (Fig. 2c), and RLIM protein was undetectable in somatic tissues (Fig. 2d). Moreover, in matings using cKO/Y-SC males we never observed the transmission of an unrecombined floxed allele (data not shown). Consistent with published data^{14,15}, these results indicate complete penetrance of the SC-mediated cKO. Analysing embryos from *fl* \times WT/Y-SC crosses, the recombination of the floxed allele was robustly detectable using PCR in blastocysts as early as E3.75 (Fig. 2e), well before initiation of rXCI.

To test whether an essential function of RLIM in rXCI might be masked by the C57BL/6 mouse background, we generated first filial generation (F1) parents in mixed C57BL/6–SV129 and C57BL/6–FVB hybrid backgrounds suitable for generating SC-mediated cKO_m offspring. Although the numbers of cKO_m F2 pups were slightly underrepresented compared with maternal wild type (WT_m) pups, there was no difference between the numbers of male and female cKO_m pups (Extended Data Fig. 2a, b), indicating that this effect is XCI independent. Sequencing analyses of cKO_m/ Δ_p pups in mixed C57BL/6–SV129 backgrounds using 156 strain-specific single nucleotide polymorphisms (SNPs) distributed among all chromosomes revealed no general bias towards the C57BL/6 background (Extended Data Fig. 2c). Together with strain-independent low RLIM protein levels in the embryonic epiblast (Fig. 1 and Extended Data Fig. 1), these data indicate that the genetic background has little or no influence on pup numbers.

¹Program in Gene Function and Expression, University of Massachusetts Medical School (UMMS), Worcester, Massachusetts 01605, USA. ²Veterinary & Animal Sciences, University of Massachusetts Amherst, Amherst, Massachusetts 01003, USA. ³Department of Cell and Developmental Biology, UMMS, Worcester, Massachusetts 01605, USA. ⁴Ortenau Klinikum Lahr-Ettenheim, Institut für Pathologie, 77933 Lahr, Germany. ⁵Program in Molecular Medicine, UMMS, Worcester, Massachusetts 01605, USA. [†]Present address: Kunming Institute of Zoology, Chinese Academy of Science, Kunming 650223, China.

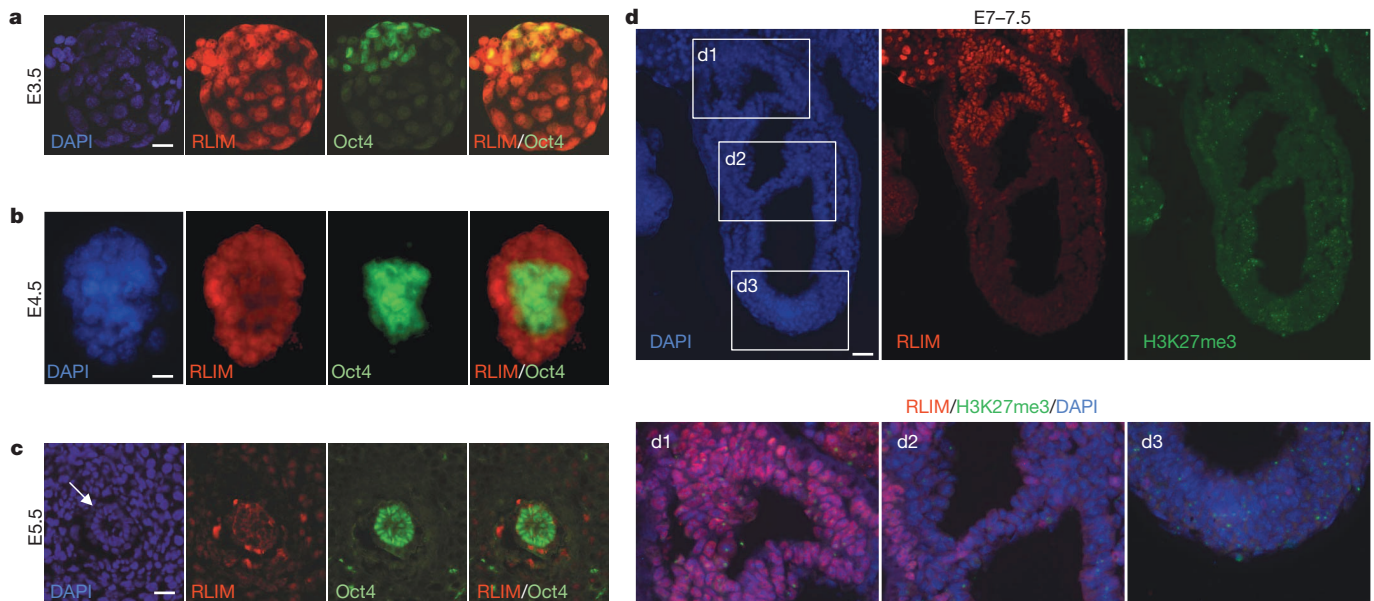


Figure 1 | Downregulation of RLIM levels in early embryonic tissues before and after implantation. **a, b,** Co-stainings of whole WT/WT pre-implantation embryos at E3.5 (**a**) and E4.5 (**b**) ($n = 5$ and 8 , respectively) using antibodies against RLIM (red) and Oct4 (green). Embryos were first photographed and then processed for genotyping. DAPI, 4',6-diamidino-2-phenylindole. **c,** Co-stainings of sections of wild-type E5.5 post-implantation embryos (arrow; $n = 10$) within placental tissues. Note the lack of RLIM staining (red) in nuclei

of Oct4-positive cells (green). **d,** Co-stainings of sections of WT/WT E7-7.5 embryos ($n = 5$) within decidual tissues using antibodies against RLIM (red) and H3K27me3 (green). Boxed regions are shown in bottom panels. d1, ectoplacental cone region; d2, amnion region; d3, embryonic epiblast region. Note the low RLIM staining intensity in cells in the amnion and epiblast regions. Scale bars, $15 \mu\text{m}$ (**a, b**), $30 \mu\text{m}$ (**c**) and $75 \mu\text{m}$ (**d**).

We next examined XCI using RNA fluorescence *in situ* hybridization (RNA-FISH) in mouse embryonic fibroblasts (MEFs) isolated from E12.5 embryos, co-hybridizing with probes specific for *Xist*¹⁷ and *Rlim*⁵. As expected, male WT/Y MEFs only showed *Rlim* transcription foci but no *Xist* clouds (Fig. 3a). In contrast, the majority of female WT/WT, cKO_m/WT, WT/cKO_p and cKO_m/Δ_p MEFs showed monoallelic expression of both genes, with *Xist* painting the inactive X and foci of *Rlim* transcription marking the active X (Xi and Xa, respectively) (Fig. 3a, b). XCI in cKO_m/Δ_p MEFs was verified by co-staining cells with antibodies directed against RLIM and H3K27me3, an XCI marker downstream of *Xist*¹⁸ (Extended Data Fig. 3). We observed a similar pattern for primary mammary epithelial cells isolated from 3-month-old adult females (Fig. 3c, d; data not shown). We also examined X skewing in MEFs isolated from female embryos with either a maternal or paternal cKO

allele, using transgenic mice containing an X-linked green fluorescent protein (*XGFP*) transgene¹⁹ on the wild-type X chromosome. MEFs of both cKO_m/*XGFP*_p and *XGFP*_m/cKO_p embryos showed approximately 1:1 ratios of GFP-positive:GFP-negative cells (Extended Data Fig. 4). These results indicate normal rXCI in somatic tissues.

Focusing our analyses on embryonic stages E5.5–E7.5, when rXCI occurs *in vivo*^{1,2}, we dissected epiblast tissue of E6–6.5 WT/WT and cKO_m/Δ_p embryos and measured mRNA levels of *Rlim* and *Xist* by quantitative PCR with reverse transcription (RT-qPCR). Although levels of *Rlim* were greatly diminished in cKO_m/Δ_p embryos, those of *Xist* were only slightly reduced (Fig. 4a). The residual levels of *Rlim* in cKO_m/Δ_p embryos are probably due to minor amounts of contaminating extra-embryonic tissue in epiblast dissections. Consistent with the kinetics of developing H3K27me3 marks on the Xi¹⁸, we first detected H3K27me3

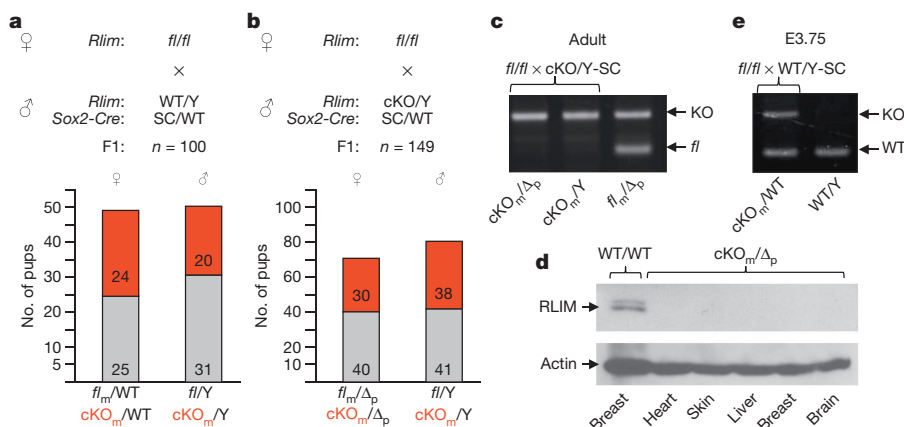


Figure 2 | *Rlim* is dispensable at post-implantation stages in female mice. **a, b,** Schematic diagram showing parental genotypes, the total number (n) of born and genotyped F1 offsprings. Number, gender and genotypes of offspring are indicated. m (maternal) and p (paternal) indicate the origin of the floxed *Rlim* allele. Maternally transmitted cKO alleles are indicated in red. **c,** High penetrance of the SC-mediated cKO in somatic tissues. PCR analysis of genomic DNA isolated from tail tissue samples using primer combinations that

detect floxed and knockout *Rlim* alleles. Parental genotypes are indicated. **d,** Lack of RLIM in somatic tissues of adult cKO_m/Δ_p females in western blots hybridized with RLIM antibodies. **e,** Recombination via paternally transmitted *Sox2-Cre* occurs in pre-implantation embryos. PCR genotyping of E3.75 blastocysts using primers that detect wild-type and knockout *Rlim* alleles. Parental genotypes are indicated.

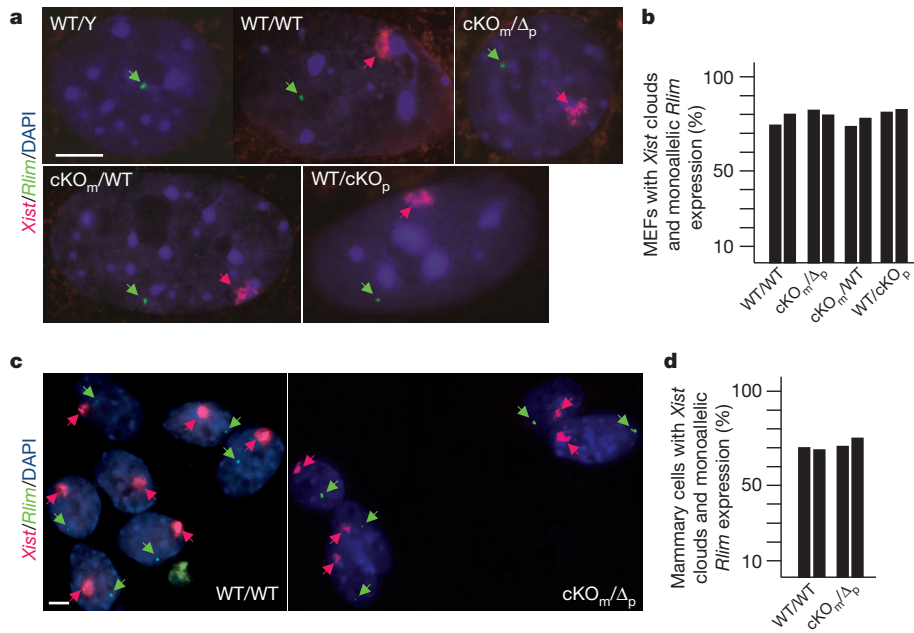


Figure 3 | Detection of *Xist* clouds in somatic tissues of cKO_m/Δ_p females. **a–d**, RNA-FISH analysis using *Xist* (red) and *Rlim* (green) as probes. **a**, RNA-FISH analysis on MEFs isolated from WT/Y, WT/WT, cKO_m/WT , WT/ cKO_p and cKO_m/Δ_p embryos at E12.5. Representative images are shown. **b**, Summary graph of RNA-FISH experiments on MEFs. Numbers of counted cells from independent biological duplicates are 51/53 (WT/WT); 79/88

foci in WT/WT embryonic tissues at stages E7–E7.5 but not in earlier post-implantation embryos (Fig. 4b; data not shown). These signals were indistinguishable from those of E7–7.5 cKO_m/Δ_p embryos, indicating similar XCI kinetics in the absence of RLIM. Moreover, cKO_m/Δ_p and WT/WT embryos were similar in size, and in immunostainings using antibodies directed against cleaved caspase 3, no increase in cells undergoing apoptosis was detected in cKO_m/Δ_p embryos (data not shown). To obtain definitive evidence that rXCI does not require the presence

(cKO_m/Δ_p); 101/83 (cKO_m/WT); 75/77 (WT/ cKO_p). **c**, RNA-FISH analysis on primary mammary epithelial cells isolated from virgin cKO_m/Δ_p adult females and WT/WT, as control. Arrows indicate specific signals. Scale bar, 3 μ m. **d**, Summary graph of RNA-FISH experiments. Numbers of counted cells from independent biological duplicates are 103/108 (WT/WT); 102/121 (cKO_m/Δ_p).

of RLIM in cells, we performed tetraploid complementation assays in which tetraploid WT/WT embryos were aggregated with low-passage *Rlim* Δ/Δ (ES-cell line IB8 or IB11), and control f_l/Δ_p (IB6) or male Δ_m/Y ES cells (and IB13), freshly isolated from blastocysts. These ES-cell lines were generated by a $f_l/\Delta_p \times \Delta/Y$ -SC cross, and are capable of developing H3K27me3 foci upon differentiation in culture, indicating XCI (Extended Data Fig. 5a, b). In tetraploid injections, the development of embryos derived from Δ/Δ ES cells was indistinguishable from

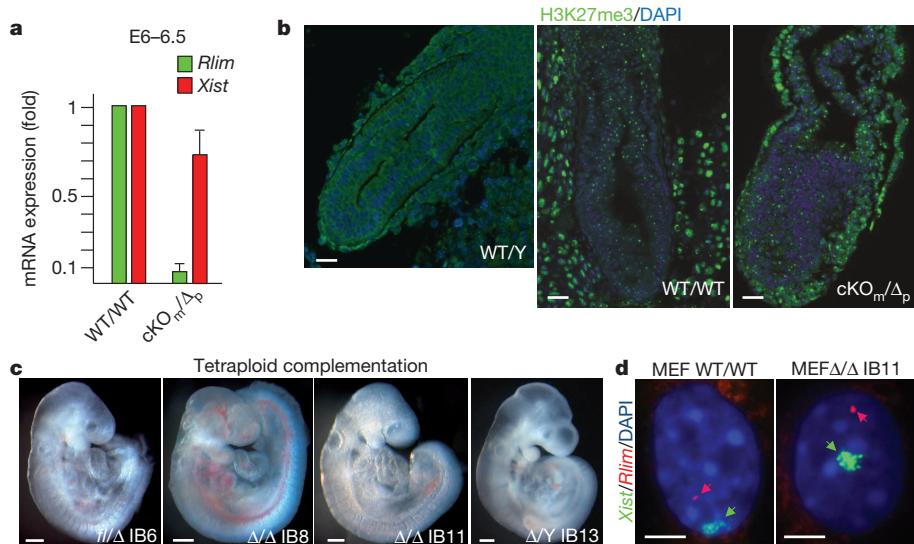


Figure 4 | Normal kinetics of XCI in female mouse embryos lacking RLIM. **a**, RT-qPCR analysis on RNA preparations prepared from epiblast dissections of E6–6.5 cKO_m/Δ_p embryos examining expression of *Rlim* and *Xist*. Values from three independent experiments were normalized against actin and are shown relative to WT/WT control. Error bars represent standard deviation. **b**, Immunohistochemical analysis of embryos sections at E7–7.5 using antibodies directed against H3K27me3. Sections of WT/Y, WT/WT and cKO_m/Δ_p embryos within decidual tissues are shown. Scale bars, 75 μ m. Images show a representative embryo of at least three obtained from independent

matings. **c**, Tetraploid complementation assay using WT/WT blastocysts complemented with low-passage Δ/Δ and, as a control, with Δ/Y and f_l/Δ ES cells. Embryos were harvested at E9–9.5 and image was recorded before genotyping. Scale bars, 0.3 mm. **d**, Δ/Δ ES cells are able to undergo XCI *in vivo*. RNA-FISH experiments on MEFs isolated from E10.5 embryos generated via tetraploid complementation using Δ/Δ ES line IB11 reveals monoallelic expression of both *Xist* (green) and *Rlim* (red). Arrows indicate specific signals. Scale bars, 3 μ m.

control ES cells at E9–9.5 (Fig. 4c), and no significant differences in complementation efficiencies between these ES-cell lines were detected (Extended Data Fig. 5c). Examining rXCI in MEFs isolated from E10.5 embryos generated from Δ/Δ IB11 cells revealed that more than 75% of Δ/Δ MEFs developed H3K27me3 foci and *Xist* clouds (Fig. 4d and Extended Data Fig. 5d, e). Combined with the presence of a single *Rlim* transcription focus in cells, this indicated that ES cells lacking RLIM are capable of undergoing XCI *in vivo*. Together, our results provide strong evidence that RLIM is dispensable for rXCI in mice. Moreover, because maternal *Rlim* is crucial for iXCI and the Δ/Δ ES-cell line IB11 is SC negative (Extended Data Fig. 5a)—therefore containing maternal and paternal Δ alleles—these data indicate that the process of iXCI is not required in pre-implantation embryos for the epiblast precursor-cell lineage that will give rise to the embryo proper.

Clearly, RLIM can serve as an activator of *Xist* in some systems^{3–5,10}. The fact that in cKO_m/ Δ_p embryos, *Xist* levels are only mildly affected and H3K27me3 signals are established with kinetics similar to WT/WT (Figs 4a, b), combined with the finding that RLIM levels are downregulated in wild-type embryos (Fig. 1, Extended Data Fig. 1), suggests that in female mouse embryos an RLIM-independent mechanism is used to activate/upregulate *Xist* transcription during rXCI. This is further underscored by our results that outgrowths of E4.5 blastocyst-stage female embryos with a maternal and paternal germline *Rlim* knockout can develop *Xist* clouds in cells of the ICM⁵. Because rXCI occurred with low frequency in the *Rlim*-knockout ES-cell system³, it seems likely that two independent mechanisms for *Xist* activation exist in mice: one that is independent of RLIM and may be inactive or downregulated in certain ES-cell systems in culture, and another that is RLIM dependent. Such a model is consistent with the presence of *Xist* activators other than RLIM, such as *Jpx* RNA^{20,21}, and would explain why RLIM is necessary for XCI in specific ES cells but not in embryos. A similar case in which ES-cell systems do not accurately reflect the *in vivo* situation has been reported for *Nanog*: although *Nanog* is required for the maintenance of pluripotency in the mouse epiblast²², it is dispensable for somatic pluripotency in ES cells²³. Moreover, a specific downregulation of RLIM in mice may also provide an explanation for reported differences in its embryonic functions, as phenotypes in morphogenesis upon *Rlim* mutation have been observed in zebrafish²⁴ but not in mice⁵.

Our results suggest important functions of parental germline imprinting on RLIM expression and the rXCI process. In addition, differences in *Rlim* mutant alleles may exert some effect on the discrepancies reported, because from the *Rlim*-knockout locus, a protein containing 83 amino acids is predicted to be expressed in mice⁵ compared to 340 amino acids in the ES-cell system^{3,4}. Thus, we cannot exclude the possibility that truncated RLIM produced from the respective *Rlim*-knockout loci may differentially influence the rXCI process. In summary, although there is no doubt about the crucial function of *Rlim* for *Xist* function during iXCI, we propose that *Rlim* is dispensable for rXCI in mice.

METHODS SUMMARY

To investigate the functions of RLIM during rXCI in mice, the conditional *Rlim* knockout⁵ was targeted to the ICM in pre-implantation blastocysts using a paternally transmitted *Sox2-Cre* transgene^{14,15} as a driver. Analyses of cells and embryos at pre- and post-implantation stages was carried out as described previously using RNA-FISH²⁵, immunological staining methods^{5,26}, western blotting²⁷ and RT-qPCR⁵. *Rlim* Δ/Δ ES-cell lines were newly generated²⁸ and used for tetraploid injections. MEFs were prepared as described²⁹.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January 2013; accepted 26 March 2014.

Published online 25 May 2014.

- Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Rev. Genet.* **12**, 429–442 (2011).

- Lee, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nature Rev. Mol. Cell Biol.* **12**, 815–826 (2011).
- Barakat, T. S. *et al.* RNF12 activates *Xist* and is essential for X chromosome inactivation. *PLoS Genet.* **7**, e1002001 (2011).
- Jonkers, I. *et al.* RNF12 is an X-encoded dose-dependent activator of X chromosome inactivation. *Cell* **139**, 999–1011 (2009).
- Shin, J. *et al.* Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature* **467**, 977–981 (2010).
- Bach, I. *et al.* RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex. *Nature Genet.* **22**, 394–399 (1999).
- Ostendorff, H. P. *et al.* Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors. *Nature* **416**, 99–103 (2002).
- Jiao, B. *et al.* Paternal RLIM/Rnf12 is a survival factor for milk-producing alveolar cells. *Cell* **149**, 630–641 (2012).
- Jiao, B. *et al.* Functional activity of RLIM/Rnf12 is regulated by phosphorylation-dependent nucleocytoplasmic shuttling. *Mol. Biol. Cell* **24**, 3085–3096 (2013).
- Gontan, C. *et al.* RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**, 386–390 (2012).
- Dupont, C. & Gribnau, J. Different flavors of X-chromosome inactivation in mammals. *Curr. Opin. Cell Biol.* **25**, 314–321 (2013).
- Schulz, E. G. & Heard, E. Role and control of X chromosome dosage in mammalian development. *Curr. Opin. Genet. Dev.* **23**, 109–115 (2013).
- Ostendorff, H. P. *et al.* Dynamic expression of LIM cofactors in the developing mouse neural tube. *Dev. Dyn.* **235**, 786–791 (2006).
- Hayashi, S., Lewis, P., Pevny, L. & McMahon, A. P. Efficient gene modulation in mouse epiblast using a *Sox2Cre* transgenic mouse strain. *Mech. Dev.* **119** (suppl. 1), S97–S101 (2002).
- Hayashi, S., Tenzen, T. & McMahon, A. P. Maternal inheritance of Cre activity in a *Sox2Cre* deleter strain. *Genesis* **37**, 51–53 (2003).
- Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
- Panning, B. X inactivation in mouse ES cells: histone modifications and FISH. *Methods Enzymol.* **376**, 419–428 (2003).
- Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
- Hadjantonakis, A. K., Cox, L. L., Tam, P. P. & Nagy, A. An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. *Genesis* **29**, 133–140 (2001).
- Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, *Jpx*, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390–403 (2010).
- Sun, S. *et al.* *Jpx* RNA activates *Xist* by evicting CTCF. *Cell* **153**, 1537–1551 (2013).
- Mitsui, K. *et al.* The homeoprotein *Nanog* is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Chambers, I. *et al.* *Nanog* safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Zhang, L. *et al.* RNF12 controls embryonic stem cell fate and morphogenesis in zebrafish embryos by targeting *Smad7* for degradation. *Mol. Cell* **46**, 650–661 (2012).
- Hall, L. L. *et al.* An ectopic human *XIST* gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl Acad. Sci. USA* **99**, 8677–8682 (2002).
- Griffith, G. J. *et al.* Yin-Yang1 is required in the mammalian oocyte for follicle expansion. *Biol. Reprod.* **84**, 654–663 (2011).
- Tursun, B. *et al.* The ubiquitin ligase Rnf6 regulates local LIM kinase 1 levels in axonal growth cones. *Genes Dev.* **19**, 2307–2319 (2005).
- Robertson, E. J. *Teratocarcinomas and Embryonic Stem Cells—A Practical Approach* Ch. 4 71–112 (IRL, 1987).
- Nagy, A. *et al.* *Manipulating the Mouse Embryo—A Laboratory Manual* 3rd edn, Ch. 8 359–397 (Cold Spring Harbor, 2003).

Acknowledgements We are grateful to H. Ma for help in the mouse facility, M. Keeler for culturing ES cells, T. Fazio, N. Lawson and Y. Yoon for advice, and T. Fazio for helpful discussions and reading of the manuscript. I.B. is a member of the University of Massachusetts Diabetes-Endocrinology Research Center (DK32520). This work was supported by National Institutes of Health grants CA131158 to I.B., CA077735 to S.N.J. and GM053234 to J.B.L.

Author Contributions J.S. and I.B. conceived and designed the experiments. M.Bo. generated the floxed *Rlim* mice. J.G. and S.N.J. generated the ES-cell lines and performed the tetraploid injections. M.Bo. and J.B.L. carried out RNA-FISH experiments and M.C.W., C.M. and J.M. performed immunohistochemistry on early embryos. All authors analysed the data. I.B. wrote the manuscript with input from J.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.B. (Ingolf.bach@umassmed.edu).

The sonic hedgehog factor GLI1 imparts drug resistance through inducible glucuronidation

Hiba Ahmad Zahreddine¹, Biljana Culjkovic-Kraljacic¹, Sarit Assouline², Patrick Gendron¹, Andrea A. Romeo³, Stephen J. Morris³, Gregory Cormack¹, James B. Jaquith^{3,†}, Leandro Cerchietti⁴, Eftihia Cocolakis², Abdellatif Amri¹, Julie Bergeron⁵, Brian Leber⁶, Michael W. Becker⁷, Shanshan Pei⁸, Craig T. Jordan⁸, Wilson H. Miller Jr² & Katherine L. Borden¹

Drug resistance is a major hurdle in oncology. Responses of acute myeloid leukaemia (AML) patients to cytarabine (Ara-C)-based therapies are often short lived with a median overall survival of months^{1–4}. Therapies are under development to improve outcomes and include targeting the eukaryotic translation initiation factor (eIF4E) with its inhibitor ribavirin^{5–7}. In a Phase II clinical trial in poor prognosis AML⁵, ribavirin monotherapy yielded promising responses including remissions; however, all patients relapsed. Here we identify a novel form of drug resistance to ribavirin and Ara-C. We observe that the sonic hedgehog transcription factor glioma-associated protein 1 (GLI1) and the UDP glucuronosyltransferase (UGT1A) family of enzymes are elevated in resistant cells. UGT1As add glucuronic acid to many drugs, modifying their activity in diverse tissues⁸. GLI1 alone is sufficient to drive UGT1A-dependent glucuronidation of ribavirin and Ara-C, and thus drug resistance. Resistance is overcome by genetic or pharmacological inhibition of GLI1, revealing a potential strategy to overcome drug resistance in some patients.

To better understand the molecular basis for relapse in our clinical trial (Fig. 1a and Extended Data Fig. 1), we generated resistant cells using head and neck carcinoma FaDu and AML-M5 THP-1 cells, both of which have highly elevated eIF4E and concomitant ribavirin sensitivity^{6,7}. Two forms of drug resistance emerged, characterized by unimpaired growth in clinically achievable ribavirin concentrations and a loss of eIF4E targeting (Fig. 1b, c and Extended Data Fig. 2a–c). Type I resistant cells (FRI, THPA, THPB) had severely impaired drug uptake, whereas type II resistant cells (FRII) did not (Fig. 1d and Extended Data Fig. 2d). Type I resistance was characterized by substantial reduction of adenosine kinase (ADK) (Extended Data Fig. 3a, b). ADK acts in cellular retention of ribavirin, as unphosphorylated ribavirin is readily exported by the nucleoside transporter ENT1 (also known as SLC29A1; refs 9, 10). Indeed, ADK reduction alone imparts ribavirin resistance (Extended Data Fig. 3c, d). Analysis of our patient pool indicated that only two patients had features consistent with type I resistance (Extended Data Fig. 3e, f).

Therefore, we investigated type II resistance. Given that eIF4E was not mutated and was functional in FRII cells (Extended Data Fig. 2e–g), we examined whether the ribavirin–eIF4E interaction was disrupted, by assessing the ability of eIF4E to immunoprecipitate ³H-ribavirin (Fig. 1e). Although ³H-ribavirin is enriched ~sixfold in the eIF4E-immunoprecipitated fraction in parental cells, this interaction is lost in FRII cells, despite normal uptake and functional eIF4E.

RNA sequencing (RNA-seq) analysis revealed that 30 transcripts were differentially expressed in FRII cells, including *GLI1* messenger RNA, by 21-fold (Extended Data Table 1). Consistently, GLI1 protein levels were highly elevated, as was GLI1's target SNAIL (Fig. 1f). We investigated the clinical relevance of this elevation in our patients treated with ribavirin monotherapy. At relapse, leukaemic blasts had elevated *GLI1* mRNA

levels for 9 out of 9 patients examined (up to tenfold baseline) (Fig. 2a). For instance, *GLI1* mRNA and protein levels were elevated at relapse (for example, patients 8 (complete remission (CR)), 11 (partial remission (PR)) and 17 (blast response (BR); Fig. 2a and Extended Data Fig. 4; confocal microscopy was used owing to limited material at response). Interestingly, GLI1 levels in patient 17 decreased during response and re-emerged at relapse. Patients 9, 13 and 19, who did not respond clinically or molecularly, had highly elevated GLI1 levels before treatment,

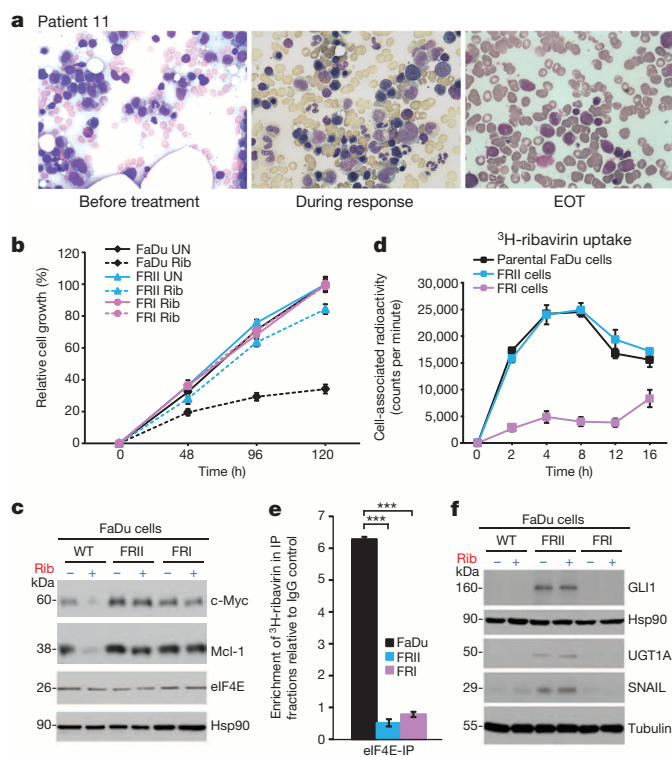


Figure 1 | Ribavirin resistance in patients and cell lines. **a**, Bone marrow biopsies for patient 11 (Wright–Giemsa staining, $\times 50$ magnification). EOT, end of treatment. **b**, Cell growth in untreated (UN) or treated with 20 μ M ribavirin (Rib), which is clinically achievable⁵. **c**, Western blots of eIF4E targets Mcl-1 and c-Myc (also known as MCL1 and MYC, respectively). **d**, ³H-ribavirin uptake. **e**, eIF4E co-immunoprecipitation (IP) with ³H-ribavirin relative to IgG. Loading is in Extended Data Fig. 6. **f**, Western blots of GLI1 and UGT1A. For **b**, **d** and **e**, error bars are means \pm standard deviations (s.d.). *** $P < 0.001$ (Student's t -test). For all panels experiments were completed in triplicate, at least three independent times. Hsp90 and tubulin provide loading controls.

¹Institute for Research in Immunology and Cancer and Department of Pathology and Cell Biology, Université de Montréal, P.O. Box 6128, Downtown Station, Montréal, Québec H3C 3J7, Canada. ²Segal Cancer Centre, Lady Davis Institute, Jewish General Hospital, McGill University, 3755 Côte-Ste-Catherine Road, Montréal, Québec H3T 1E2, Canada. ³Pharmascience Inc., 6111 Royalmount Avenue, Montréal, Québec H4P 2T4, Canada. ⁴Division of Hematology and Oncology, Department of Medicine, Weill Cornell Medical College, Cornell University, 1305 York Avenue, New York, New York 10021, USA. ⁵Hôpital Maisonneuve-Rosemont, 5415 Boulevard de l'Assomption, Montréal, Québec H1T 2M4, Canada. ⁶McMaster University/Hamilton Health Sciences, 237 Barton Street East, Hamilton, Ontario L8L 2X2, Canada. ⁷Department of Medicine, Division of Hematology/Oncology, 601 Elmwood Avenue, University of Rochester, Rochester, New York 14627, USA. ⁸Division of Hematology, Department of Medicine, University of Colorado Denver, 13123 East 16th Avenue, Aurora, Colorado 80045, USA. [†]Present address: JAQJAM Consulting, Montreal J7V 9B6, Canada.

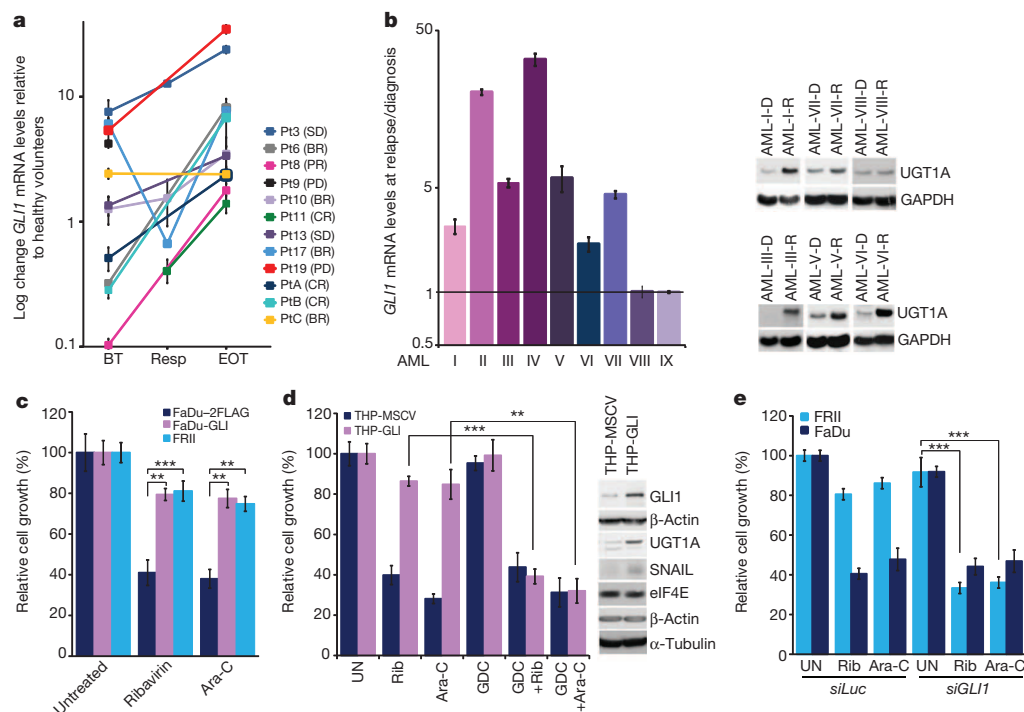


Figure 2 | GLI1 status underpins resistance. **a**, *GLI1* mRNA levels before treatment (BT), at best response (Resp) and EOT for ribavirin-treated patients/healthy volunteers. Pt, patient. **b**, *GLI1* mRNA (right) and UGT1A protein (left) for patients treated with Ara-C therapies at relapse (R)/diagnosis (D). Roman numerals denote patients treated with standard chemotherapy as opposed to ribavirin monotherapy (numbers) or ribavirin plus Ara-C (letters). **c**, **d**, Effects of GLI1 overexpression on drug sensitivity: ribavirin (20 μ M),

Ara-C (200 nM, clinically achievable⁷). **e**, siRNA to *GLI1* restores drug sensitivity. *siLuc*, nonspecific control RNA. **a**, **b**, RNA results were normalized to ubiquitin. In all panels, error bars denote mean \pm s.d. Experiments were completed in triplicate at least three independent times. ** $P < 0.01$, *** $P < 0.001$ (Student's *t*-test). Figure 4a–c shows corresponding western blots. Actin, tubulin and GAPDH provide loading controls.

relative to healthy individuals or responding patients. Moreover, in our ribavirin and low-dose Ara-C Phase I combination trial, patients A (CR, relapsed at 2 years) and B (CR, relapsed at 9 months) had increased *GLI1* levels at relapse (Fig. 2a and Extended Data Fig. 4; these patients are denoted by letters so as not to be confused with ribavirin monotherapy patients, denoted by numbers). For patient C (PR, off treatment owing to a dose-limiting toxicity), *GLI1* mRNA levels were unchanged at end of treatment (EOT), consistent with continued remission at that time. Thus, elevated *GLI1* is associated with primary and acquired resistance in both ribavirin clinical trials. We observed that type I and type II resistance coexisted in two patients (Fig. 2 and Extended Data Fig. 3e, f). Whether this occurs in the same cells or whether multiple resistant populations emerge is unknown.

We extended these studies to investigate whether *GLI1* levels were elevated in patients who failed more commonly used Ara-C therapies (Fig. 2b). *GLI1* mRNA levels were substantially elevated at relapse relative to diagnosis in 7 out of 9 patients examined. Consistently, analysis of The Cancer Genome Atlas AML data set¹¹ showed that *GLI1* elevation was associated with poor prognosis (Extended Data Fig. 5). These data suggest that *GLI1* overexpression contributed to drug resistance and clinical relapse in some AML patients.

Next, we investigated whether *GLI1* alone imparted drug resistance by generating FaDu and THP1 cells stably overexpressing *GLI1* (FaDu-GLI, THP1-GLI; Fig. 2c, d). Growth of *GLI1*-overexpressing cells was not substantially affected by levels of ribavirin or Ara-C that impaired growth of controls. Further, *GLI1* knockdown re-sensitized FRII cells to ribavirin and Ara-C without affecting growth in the absence of either drug (Fig. 2e).

To identify strategies to restore drug sensitivity, we used a clinically approved inhibitor of sonic hedgehog signalling upstream of *GLI1*, GDC-0449 (ref. 12) (Figs 2d and 3a). FRII cells were pre-treated with 200 nM GDC-0449 (which is clinically achievable¹³) and subsequently, 20 μ M

ribavirin. Notably, GDC-0449 treatment followed by ribavirin led to $\sim 60\%$ reduction in growth relative to untreated FRII cells. GDC-0449 treatment alone did not substantially affect growth in either cell line. Importantly, GDC-0449 treatment also restored sensitivity to clinically relevant Ara-C levels (200 nM). Furthermore, GDC-0449 treatment of FaDu-GLI and THP-GLI cells re-sensitized these to ribavirin and Ara-C (Figs 2d and 3a). Finally, a direct inhibitor of *GLI1*, GANT61 (ref. 14), paralleled the effects of GDC-0449 (Extended Data Fig. 6a). Thus, type II resistance is reversed by pharmacological inhibition of the sonic hedgehog pathway.

To better understand the molecular basis for resistance, we monitored the ability of eIF4E to immunoprecipitate ³H-ribavirin as a function of *GLI1* status (Fig. 3b and Extended Data Fig. 6b–d). Although eIF4E–ribavirin complexes were readily detected in controls, they were absent in *GLI1*-overexpressing cells (Fig. 3b). Conversely, GDC-0449 treatment or *GLI1* knockdown in FRII cells restored ribavirin–eIF4E complexes to control levels (Fig. 3b). Thus, there is a clear correlation between *GLI1* elevation, reduction in eIF4E–ribavirin complexes, and resistance.

Given that resistant cells did not form ribavirin–eIF4E complexes but retained active eIF4E (Figs 1e, 3b and Extended Data Fig. 2e–g), we proposed that ribavirin, and possibly Ara-C, underwent some form of *GLI1*-dependent modification. The drug-metabolizing UGT1A enzymes had elevated protein levels in FRII cells, thereby suggesting a resistance mechanism (Figs 1f and 4a–c). This was also the case for FaDu-GLI and THP-GLI cells, relative to vector controls (Figs 2d and 4b). Significantly, *GLI1* knockdown or GDC-0449 treatment reduced UGT1A protein levels (Fig. 4a, c), confirming the correlation between *GLI1* and UGT1A protein expression. Note that *GLI1* does not increase mRNA levels but rather the protein stability of UGT1As (Extended Data Fig. 6e–h).

To determine the clinical relevance of these observations, we examined UGT1A protein levels during response and relapse using confocal microscopy (Extended Data Fig. 4). We observed UGT1A elevation upon

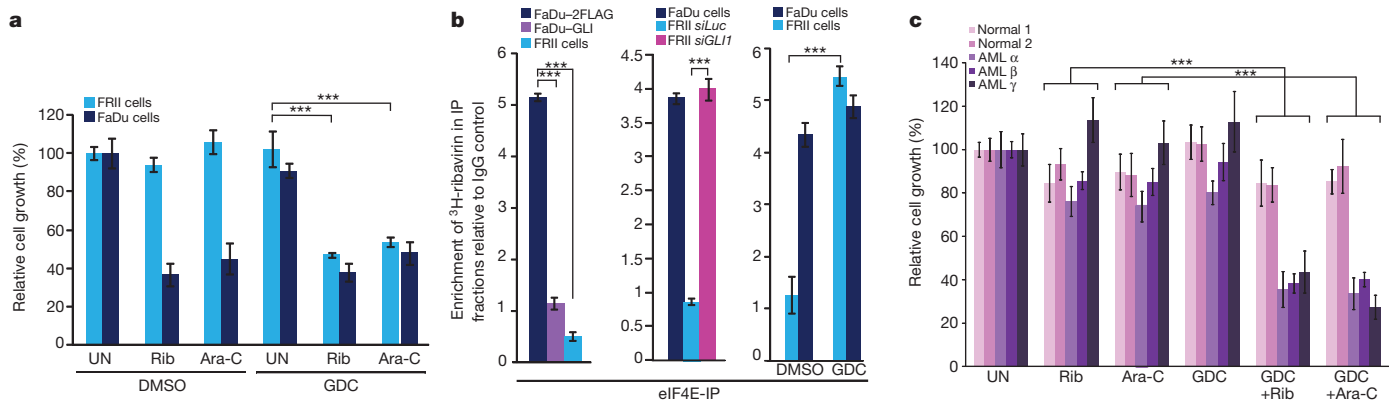


Figure 3 | Targeting GLI1 activity. **a**, FRII cell growth \pm GDC-0449 (GDC) and ribavirin or Ara-C. DMSO, dimethylsulphoxide. **b**, The presence of the ribavirin-eIF4E complex is measured as a function of GLI1 status. Ribavirin-eIF4E complexes are detected using ³H-ribavirin and eIF4E immunoprecipitation relative to IgG controls: FaDu-GLI cells, *siGLI1*, GDC-0449 treatment. See Extended Data Fig. 6 for immunoprecipitation

controls. **c**, Methylcellulose colony growth assays in specimens from M4/M5 AML patients or healthy volunteers (bone marrow). For all panels error bars are mean \pm s.d. Results are representative of at least three independent experiments in triplicate (**a**, **b**) or at least two (**c**) in replicates of 5. ***P* < 0.01, ****P* < 0.001 (Student's *t*-test).

relapse; that is, patients 11 (CR), 8 (PR) and 17 (BR) in the ribavirin monotherapy trial and in patients A (CR) and B (CR) in the combination trial. Patient C (PR) had no change in UGT1A levels at EOT, consistent with still being in remission. In patients treated with standard Ara-C therapies, UGT1A protein levels were elevated in 6 out of 7 specimens at relapse relative to diagnosis, and this occurred in the patients with concomitant elevated GLI1 (Fig. 2b). There was insufficient material for protein analysis of the remaining two specimens.

Next, we used mass spectrometry (MS) to determine whether ribavirin and Ara-C were glucuronidated in resistant cells (Fig. 4d–i and Extended Data Fig. 7). Metabolites were isolated, subjected to hydrophilic chromatography and detected by electrospray ionization-MS. In parental cells, ribavirin diphosphate (RDP) is the major peak (Fig. 4e, l).

In FRII cells, a new peak emerged with a mass consistent with the ribavirin-glucuronide (Fig. 4d). Using collision-induced ion fragmentation, we observed the triazole moiety of ribavirin as a major fragment supporting this as a site of glucuronidation (Fig. 4j red arrow, and Extended Data Fig. 7a). Relative peak intensities suggest that there is more ribavirin-glucuronide than RDP (Fig. 4d). Notably, GDC-0449 treatments eliminated ribavirin glucuronidation in FRII cells (Fig. 4f). GLI1 overexpression in parental cells led to formation of ribavirin-glucuronides (Fig. 4h). *In vitro* glucuronidation studies indicated that specific UGT1As are likely to be important to this process, as is ribavirin phosphorylation (Extended Data Fig. 7). Moreover, we observe Ara-C-glucuronides in FRII but not parental cells, and this modification was lost upon GDC-0449 treatment (Extended Data Fig. 7e, f). Thus, Ara-C and ribavirin

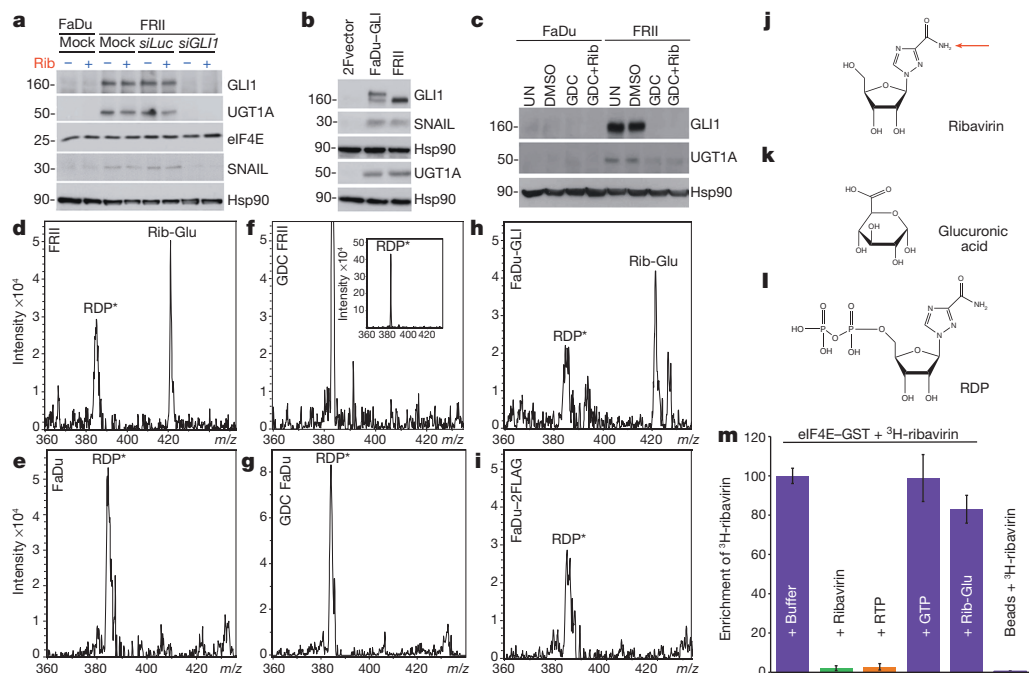


Figure 4 | Link between GLI1, UGT1A and drug glucuronidation. **a–c**, Western blots as a function of GLI1. Hsp90 provides a loading control. 2F, 2FLAG. **d–i**, MS/MS extracted ion chromatograms are shown. Rib-Glu, ribavirin glucuronide; RDP*, ribavirin diphosphate minus H₂O. Fragmentation of Rib-Glu peak confirms its identity (Extended Data Fig. 7a). **j**, Ribavirin, the best-supported glucuronidation site indicated by red

arrow. **k**, Glucuronic acid. **l**, RDP. **m**, ³H-ribavirin-eIF4E-GST complexes competed by equimolar amounts of compounds. ³H-ribavirin did not bind beads alone. See Extended Data Fig. 7d for loading. Error bars are averages \pm s.d. Experiments were carried out in triplicate, at least three independent times.

glucuronidation were GLI1-dependent, and elimination of the glucuronides by GLI1 inhibition correlated with restored drug sensitivity.

We examined the capacity of ribavirin-glucuronides to bind eIF4E. Ribavirin-glucuronide was isolated by hydrophilic chromatography and confirmed by MS/MS (Extended Data Fig. 7c). Using eIF4E–glutathione S-transferase (GST) immobilized on glutathione agarose, we observed that non-radioactive ribavirin or ribavirin triphosphate (RTP) compete for ^3H -ribavirin–eIF4E complexes, whereas neither the negative control GTP nor the ribavirin-glucuronide did so (Fig. 4m). Thus, ribavirin glucuronidation impairs its interaction with eIF4E, underpinning resistance.

To further explore the effects of GLI1 inhibition on drug sensitivity, we monitored colony growth of primary AML specimens as a function of GDC-0449 treatment (Fig. 3c). Specimens were selected from patients that had previously failed induction chemotherapy. We observed that although GDC-0449 has little effect on colony growth alone, it strongly potentiated the effects of Ara-C and ribavirin, presumably by elimination of the respective glucuronides. By contrast, we observed little effect in specimens from healthy volunteers, consistent with our results in control cells.

Several factors probably contribute to GLI1 elevation in FRII cells, including reduced patched 1 levels, but not altered promoter methylation or modified hedgehog ligand levels (Extended Data Fig. 8). Glucuronidation is typically perceived as a detoxification pathway but does not always enhance drug excretion¹⁵. Similar to our findings with ribavirin and eIF4E, testosterone glucuronidation modifies its targets rather than its efflux¹⁵. Our findings reveal a role for GLI1 in drug metabolism and resistance. Here, GLI1 inhibition could restore drug sensitivity and thereby provide therapeutic benefit.

METHODS SUMMARY

Ribavirin-resistant cell lines were selected on the basis of prolonged ribavirin exposure, routinely tested for resistance, which was retained even after 6 months in the absence of ribavirin. In the absence of ribavirin, cells grew with indistinguishable doubling times (Extended Data Fig. 2). Drug treatments and cell viability assays were carried out as described¹⁶ using Trypan Blue or in parallel, Cell Counting Kit-8. For ^3H -ribavirin immunoprecipitation, cells were incubated with $0.66\ \mu\text{M}$ ^3H -ribavirin for 8 h at 37°C , dounce homogenized, lysates pre-cleared and immunoprecipitated with $8\ \mu\text{g}$ mouse IgG or monoclonal mouse anti-eIF4E antibodies coupled to Protein A/G PLUS-Agarose overnight at 4°C . Beads were washed, subjected to scintillation counting or western analysis.

Quantitative PCR analysis used the relative standard curve method⁵. Western analysis was described⁶ with a modified lysis buffer (40 mM HEPES, pH 7.5, 120 mM NaCl, 1 mM EDTA, 10 mM β -glycerophosphate, 50 mM NaF, $0.5\ \mu\text{M}$ NaVO_3 , and 1% (v/v) Triton X-100 with protease inhibitors). For mass spectroscopy, cells were treated with ribavirin or Ara-C, mixed with an equal volume of methanol, flash frozen in liquid nitrogen, stored at -80°C overnight, thawed, centrifuged at 10,000 r.p.m. for 10 min and subjected to hydrophilic chromatography (Inertsil HILIC) using an Agilent 1100 HPLC coupled to an Agilent MSD Trap SL, with an ESI source in positive ion mode. Chromatography used 2 mM ammonium formate, pH 3.2 (solvent A) and 100% acetonitrile (solvent B). For ^3H -ribavirin pulldown, eIF4E–GST immobilized on glutathione beads was incubated with $250\ \text{nM}$ ^3H -ribavirin and the same concentration of purified Rib-Glu, ^1H -ribavirin, RTP, GTP or buffer for 30 min in 100 mM sodium phosphate, pH 7.5, 300 mM NaCl, 0.015% NP40, 10 μM BSA with protease inhibitors. Samples were eluted in Laemmli; radioactivity measured by scintillation counting. Colony formation assays in AML specimens were as described⁷.

Clinical Trials were approved by each institution ethics board and by Health Canada. Written informed consent was obtained according to the Helsinki protocol. Clinical trials were registered with ClinicalTrials.gov NCT00559091 and NCT01056523.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 May 2013; accepted 26 March 2014.

Published online 28 May 2014.

1. Deschler, B., de Witte, T., Mertelsmann, R. & Lubbert, M. Treatment decision-making for older patients with high-risk myelodysplastic syndrome or acute myeloid leukemia: problems and approaches. *Haematologica* **91**, 1513–1522 (2006).
2. Deschler, B. & Lubbert, M. Acute myeloid leukemia: epidemiology and etiology. *Cancer* **107**, 2099–2107 (2006).
3. Pulsoni, A. et al. Survival of elderly patients with acute myeloid leukemia. *Haematologica* **89**, 296–302 (2004).
4. Hiddemann, W. et al. Management of acute myeloid leukemia in elderly patients. *J. Clin. Oncol.* **17**, 3569–3576 (1999).
5. Assouline, S. et al. Molecular targeting of the oncogene eIF4E in acute myeloid leukemia (AML): a proof-of-principle clinical trial with ribavirin. *Blood* **114**, 257–260 (2009).
6. Kentsis, A., Topisirovic, I., Culjkovic, B., Shao, L. & Borden, K. L. Ribavirin suppresses eIF4E-mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap. *Proc. Natl Acad. Sci. USA* **101**, 18105–18110 (2004).
7. Kraljacic, B. C., Arguello, M., Amri, A., Cormack, G. & Borden, K. Inhibition of eIF4E with ribavirin cooperates with common chemotherapies in primary acute myeloid leukemia specimens. *Leukemia* **25**, 1197–1200 (2011).
8. Tukey, R. H. & Strassburg, C. P. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.* **40**, 581–616 (2000).
9. Jarvis, S. M., Thorn, J. A. & Glue, P. Ribavirin uptake by human erythrocytes and the involvement of nitrobenzylthioinosine-sensitive (es)-nucleoside transporters. *Br. J. Pharmacol.* **123**, 1587–1592 (1998).
10. Willis, R. C., Carson, D. A. & Seegmiller, J. E. Adenosine kinase initiates the major route of ribavirin activation in a cultured human cell line. *Proc. Natl Acad. Sci. USA* **75**, 3042–3044 (1978).
11. Cancer Genome Atlas Research. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
12. Mas, C. & Ruiz i Altaba, A. Small molecule modulation of HH-Gli signaling: current leads, trials and tribulations. *Biochem. Pharmacol.* **80**, 712–723 (2010).
13. LoRusso, P. M. et al. Pharmacokinetic dose-scheduling study of hedgehog pathway inhibitor vismodegib (GDC-0449) in patients with locally advanced or metastatic solid tumors. *Clin. Cancer Res.* **17**, 5774–5782 (2011).
14. Lauth, M., Bergstrom, A., Shimokawa, T. & Toftgard, R. Inhibition of Gli-mediated transcription and tumor cell growth by small-molecule antagonists. *Proc. Natl Acad. Sci. USA* **104**, 8455–8460 (2007).
15. Dutton, G. *Glucuronidation of Drugs and Other Compounds* Ch. 1 (CRC, 1980).
16. Olah, E. et al. Molecular mechanisms in the antiproliferative action of taxol and tiazofurin. *Anticancer Res.* **16**, 2469–2477 (1996).

Acknowledgements We are grateful for reading of the manuscript by A. Kentsis, M. Osborne, F. Pettersson, S. del Rincon and C. Perreault. We are grateful for the use of Pharmascience facilities thanks to M. Goodman and D. Goodman, for specimens from the Banques de Cellules Leucémiques Du Québec (BCLQ), and for technical discussions with M. Ouellet. We thank nursing and support staff at all the clinical sites, and A. Ramteke for technical support. K.L.B.B. is supported by funds from the National Institutes of Health (R01 80728 and 98571), IRICoR and Translational Research Program grants from the Leukemia and Lymphoma Society USA. K.L.B.B. holds a Canada Research Chair and H.A.Z. holds a Cole Foundation Fellowship and a CNRS Lebanon Fellowship.

Author Contributions H.A.Z. and B.C.-K. designed and performed experiments, analysed data and wrote the manuscript; L.C., A.A. and G.C. performed experiments; P.G., S.P. and C.T.J. analysed data; A.A.R. performed experiments and analysed data; J.B.J. and S.J.M. designed experiments; S.A. was medical monitor for the trial, treated patients and analysed clinical data; E.C. coordinated clinical trials and analysed clinical data; B.L., J.B. and M.W.B. treated patients, provided specimens and analysed clinical data; W.H.M. designed experiments, analysed data and edited the manuscript; K.L.B.B. designed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to K.L.B.B. (katherine.borden@umontreal.ca).

Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones

Juan Pablo Maianti¹, Amanda McFedries¹, Zachariah H. Foda², Ralph E. Kleiner¹, Xiu Quan Du³, Malcolm A. Leissring⁴, Wei-Jen Tang⁵, Maureen J. Charron³, Markus A. Seeliger², Alan Saghatelian¹ & David R. Liu^{1,6}

Despite decades of speculation that inhibiting endogenous insulin degradation might treat type-2 diabetes^{1,2}, and the identification of *IDE* (insulin-degrading enzyme) as a diabetes susceptibility gene^{3,4}, the relationship between the activity of the zinc metalloprotein IDE and glucose homeostasis remains unclear. Although *Ide*^{-/-} mice have elevated insulin levels, they exhibit impaired, rather than improved, glucose tolerance that may arise from compensatory insulin signalling dysfunction^{5,6}. IDE inhibitors that are active *in vivo* are therefore needed to elucidate IDE's physiological roles and to determine its potential to serve as a target for the treatment of diabetes. Here we report the discovery of a physiologically active IDE inhibitor identified from a DNA-templated macrocycle library. An X-ray structure of the macrocycle bound to IDE reveals that it engages a binding pocket away from the catalytic site, which explains its remarkable selectivity. Treatment of lean and obese mice with this inhibitor shows that IDE regulates the abundance and signalling of glucagon and amylin, in addition to that of insulin. Under physiological conditions that augment insulin and amylin levels, such as oral glucose administration, acute IDE inhibition leads to substantially improved glucose tolerance and slower gastric emptying. These findings demonstrate the feasibility of modulating IDE activity as a new therapeutic strategy to treat type-2 diabetes and expand our understanding of the roles of IDE in glucose and hormone regulation.

To discover small-molecule modulators of IDE, we performed *in vitro* selections on a DNA-templated library of 13,824 synthetic macrocycles^{7,8} for the ability to bind immobilized mouse IDE, resulting in six candidate IDE-binding molecules (Extended Data Fig. 1). The 20-membered macrocycle **6b** (Fig. 1a, half-maximum inhibitory concentration IC₅₀ = 60 nM) potentially inhibited IDE activity in three complementary assays (Extended Data Fig. 2)⁹. We synthesized and biochemically assayed 30 analogues of **6b** in which each building block was systematically varied to elucidate the structural and stereochemical requirements (Extended Data Fig. 1), and based on the results we identified the inhibitor **6bK** (IC₅₀ = 50 nM, Fig. 1b) as an ideal candidate for *in vivo* studies.

The selectivity of **6bK** *in vitro* was ≥1,000-fold for inhibition of IDE over all other metalloproteases tested, a substantial improvement over the previously reported substrate mimetic hydroxamic acid inhibitor **Ii1**¹⁰ (Fig. 1d, e). The selectivity of **6bK**, coupled with its ability to inhibit IDE in a synergistic, rather than competitive, manner with **Ii1** (Extended Data Fig. 2), led us to speculate that the macrocycle engages a binding site distinct from the enzyme's catalytic site (Supplementary Discussion).

We determined the X-ray crystal structure of catalytically inactive cysteine-free human IDE¹¹ bound to **6b** at 2.7 Å resolution (Fig. 1g, Extended Data Fig. 3). Macrocycle **6b** occupies a binding pocket at the interface of IDE domains 1 and 2, and is positioned 11 Å away from the catalytic zinc ion (Fig. 1h). This distal binding site is a unique structural feature of IDE compared to related metalloproteases¹², and does not

overlap with the binding site of **Ii1**¹⁰. IDE mutations predicted by the structure to impede macrocycle binding led to losses of **6bK** potency (Fig. 1f), and complementary changes in **6b** analogues rescued inhibition (Supplementary Discussion, Extended Data Fig. 4). The structure predicts that by engaging this distal site the macrocycle precludes substrate binding and abrogates key interactions that are necessary to unfold peptides for cleavage (Supplementary Video)^{13,14}.

We characterized the stability, and the physicochemical and pharmacokinetic properties, of **6bK** formulated in Captisol¹⁵, a β-cyclodextrin agent used to improve delivery through intraperitoneal (i.p.) injection at 2 mg **6bK** per animal (Supplementary Discussion, Extended Data Fig. 5). The long half-life in mouse plasma (>2 h) and in circulation (>1 h) of **6bK** suggested that it was suitable for *in vivo* studies (Extended Data Fig. 5). Injection of **6bK** resulted in high levels of the inhibitor (>100-fold IC₅₀) in peripheral circulation and in the liver and kidneys, the main insulin-degrading organs. In contrast, **6bK** was undetectable in brain tissue, where IDE is known to degrade amyloid peptides⁵ (Extended Data Fig. 5), and levels of Aβ(40) and Aβ(42) peptides in mice injected with **6bK** were unchanged (Extended Data Fig. 5). Taken together, these findings suggested the viability of **6bK** as an *in vivo* IDE inhibitor.

To evaluate the ability of **6bK** to inhibit IDE activity *in vivo*, we subjected non-fasted mice to insulin tolerance tests 30 min after a single injection with **6bK** (2 mg per animal, 80 mg kg⁻¹). Mice treated with **6bK** experienced lower hypoglycaemia and higher insulin levels compared to vehicle controls (*P* < 0.01, see below and Extended Data Fig. 5). These experiments establish that a selective and physiologically stable pharmacological IDE inhibitor can augment the abundance and activity of insulin *in vivo*.

Next we determined the physiological consequences of acute IDE inhibition for glucose tolerance. We used two methods of glucose delivery (oral gavage or i.p. injection¹⁶) and two different mouse models (lean or diet-induced obese (DIO) mice^{17,18}). These four conditions were chosen to survey the role of IDE activity under a broad range of endogenous insulin levels and insulin sensitivities^{16,17}. Oral glucose administration results in greater insulin secretion compared to injected glucose delivery (Extended Data Fig. 6) that arises from the 'incretin effect'^{17,19}. DIO mice display hyperinsulinaemia and insulin resistance, and serve as a model for early type-2 diabetes in humans¹⁸.

During oral glucose tolerance tests (OGTTs, Fig. 2a and b), both lean and DIO mice treated with **6bK** displayed significantly improved glucose tolerance compared to control groups treated either with vehicle alone, or with the inactive stereoisomer **bisepi-6bK** (Figs 1c, 2a and b; Extended Data Fig. 7). Effects of similar magnitude on oral glucose tolerance in mice have been observed from several human diabetes therapeutics (Extended Data Fig. 8)^{19–21}. The vehicle and **bisepi-6bK** control groups exhibited similar blood glucose profiles, indicating that the effects of **6bK** on glucose tolerance are lost when the stereochemistry of **6bK** is altered in a way that abolishes IDE inhibition. Collectively, these

¹Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. ²Department of Pharmacological Sciences, Stony Brook University, 1 Circle Road, Stony Brook, New York 11794, USA. ³Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ⁴Institute for Memory Impairments and Neurological Disorders, University of California, Irvine, 3204 Biological Sciences III, Irvine, California 92697, USA. ⁵Ben-May Department for Cancer Research, University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. ⁶Howard Hughes Medical Institute, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA.

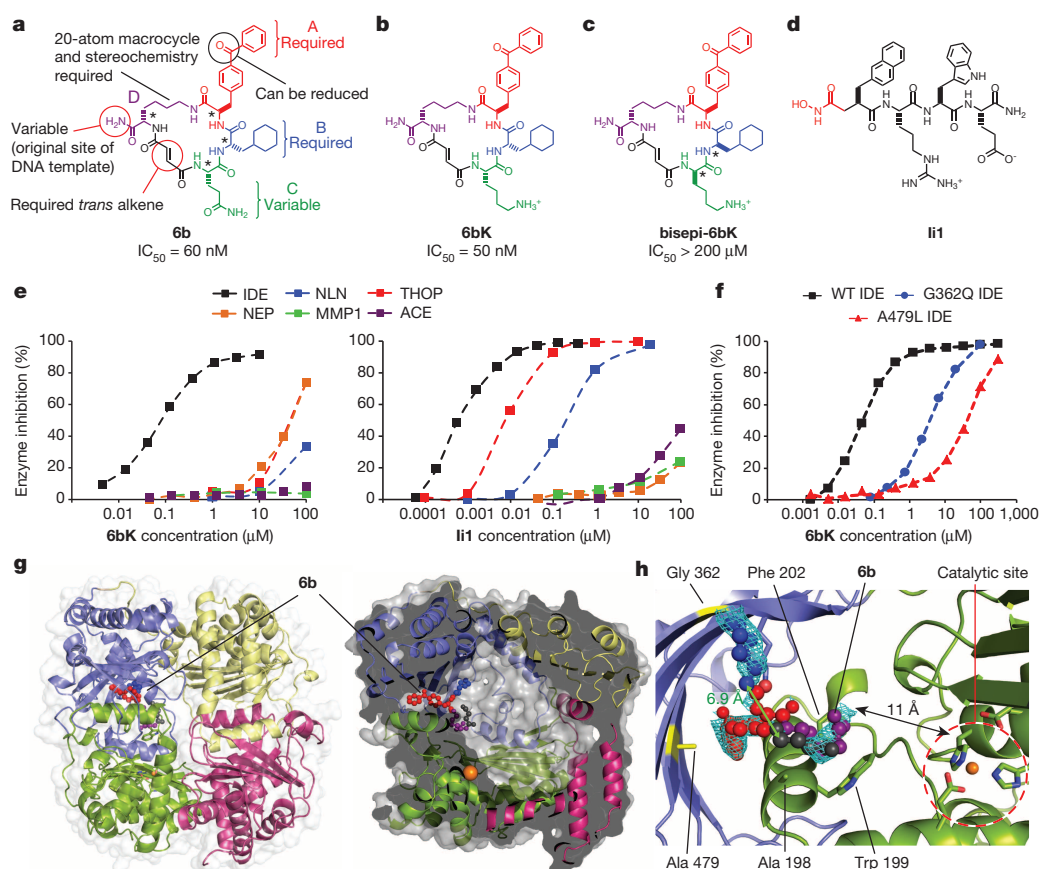


Figure 1 | Potent and highly selective macrocyclic IDE inhibitors from the *in vitro* selection of a DNA-templated macrocycle library. **a**, Structure of **6b** and summary of the requirements for IDE inhibition revealed by assaying **6b** analogues (Extended Data Fig. 1). **b**, Physiologically active IDE inhibitor **6bK**. **c**, Inactive diastereomer **bisepi-6bK**. **d**, Previously reported substrate-mimetic hydroxamic acid **Ii1**¹⁰. **e**, Selectivity analysis of macrocycle **6bK** reveals >1,000-fold selectivity for IDE (IC_{50} = 50 nM) over all other metalloproteases tested. In contrast, inhibitor **Ii1**¹⁰ inhibits IDE (IC_{50} = 0.6 nM), thimet oligopeptidase (THOP, IC_{50} = 6 nM) and neurolysin (NLN, IC_{50} = 185 nM),

but not NEP (neprilysin), MMP1 (matrix metalloproteinase, 1) or ACE (angiotensin-converting enzyme). **f**, Activity assays for wild-type or mutant human IDE variants in the presence of **6bK**. **g**, X-ray co-crystal structure of IDE bound to macrocyclic inhibitor **6b** (2.7 Å resolution, PDB 4LTE). **h**, Electron density map (composite omit map contoured at 1σ) showing the relative position of macrocycle **6b** bound 11 Å from the catalytic zinc atom. The glutamine residue and four atoms of the macrocycle backbone were unresolved. See also Extended Data Figs 2–4.

observations represent the first demonstration that transient IDE inhibition improves blood glucose tolerance¹.

We repeated the above experiments with **6bK** using i.p.-injected glucose tolerance tests (IPGTTs). In contrast to the observed improvement in oral glucose tolerance on **6bK** treatment, IDE inhibition with **6bK** followed by a glucose injection (1.5 g kg^{-1} i.p.) surprisingly resulted in impaired glucose tolerance in both lean and obese mice compared to vehicle alone or **bisepi-6bK**-treated controls (Fig. 2d and e). DIO mice treated with **6bK** followed by glucose injection displayed a biphasic response: glucose levels were lower over the initial 30 min of the IPGTT, followed by a hyperglycaemic ‘rebound’ starting 1 h after glucose injection (Fig. 2e and Extended Data Fig. 7). To further test if the effects of **6bK** are specific to its interaction with IDE, we repeated these experiments using *Ide*^{-/-} knockout mice^{5,6}. Mice lacking IDE were not affected by **6bK** treatment and exhibited OGTT and IPGTT blood glucose responses comparable to that of vehicle-treated cohorts (Fig. 2c and f), suggesting that blood glucose profile improvement during OGTT and impairment during IPGTT of **6bK**-treated wild-type mice are mediated by IDE.

The dependence of the physiological response to **6bK** on the route of glucose administration cannot be easily explained by a simple model in which IDE only degrades insulin. Instead, these results strongly suggest a role for IDE in modulating other glucose-regulating hormones *in vivo* beyond insulin. The biochemical properties of IDE and its substrate recognition mechanism^{12,13} enable this enzyme to cleave a wide range of

peptide substrates *in vitro* for which experimental validation *in vivo* has not been previously possible (Supplementary Table 1). Two glucose-regulating hormones, beyond insulin, that are potential candidates for physiological regulation by IDE are glucagon and amylin. Whereas purified IDE can cleave these two peptides *in vitro*^{22,23}, neither hormone is known to be regulated by IDE *in vivo* or *ex vivo*.

To probe the possibility that glucagon and/or amylin are regulated *in vivo* by IDE, we measured plasma levels of these hormones in DIO mice treated with **6bK** or vehicle alone following IPGTTs (Fig. 3a), and observed substantially higher levels of insulin, glucagon and amylin. Next we injected each of these three putative substrates 30 min after treatment with **6bK** or vehicle alone (Fig. 3b–d). The **6bK**-treated cohorts exhibited significantly stronger blood glucose responses to each of these hormones; mice treated with **6bK** experienced hypoglycaemia during insulin tolerance tests (Fig. 3b) and hyperglycaemia following challenges with either amylin²⁴ (Fig. 3c) or glucagon (Fig. 3d) compared to control animals (via non-physiological activation of the Cori cycle, and gluconeogenesis, respectively)^{24–26}. Moreover, in each case the plasma level of the hormone injected remained elevated at the end of the procedure in **6bK**-treated mice relative to control animals (Figs 3b–d insets). Collectively these results reveal that IDE regulates the abundance and physiological effects of glucagon and amylin, in addition to insulin.

Amylin is co-secreted with insulin, and is involved in glycaemic control by inhibiting gastric emptying, promoting satiety and antagonizing

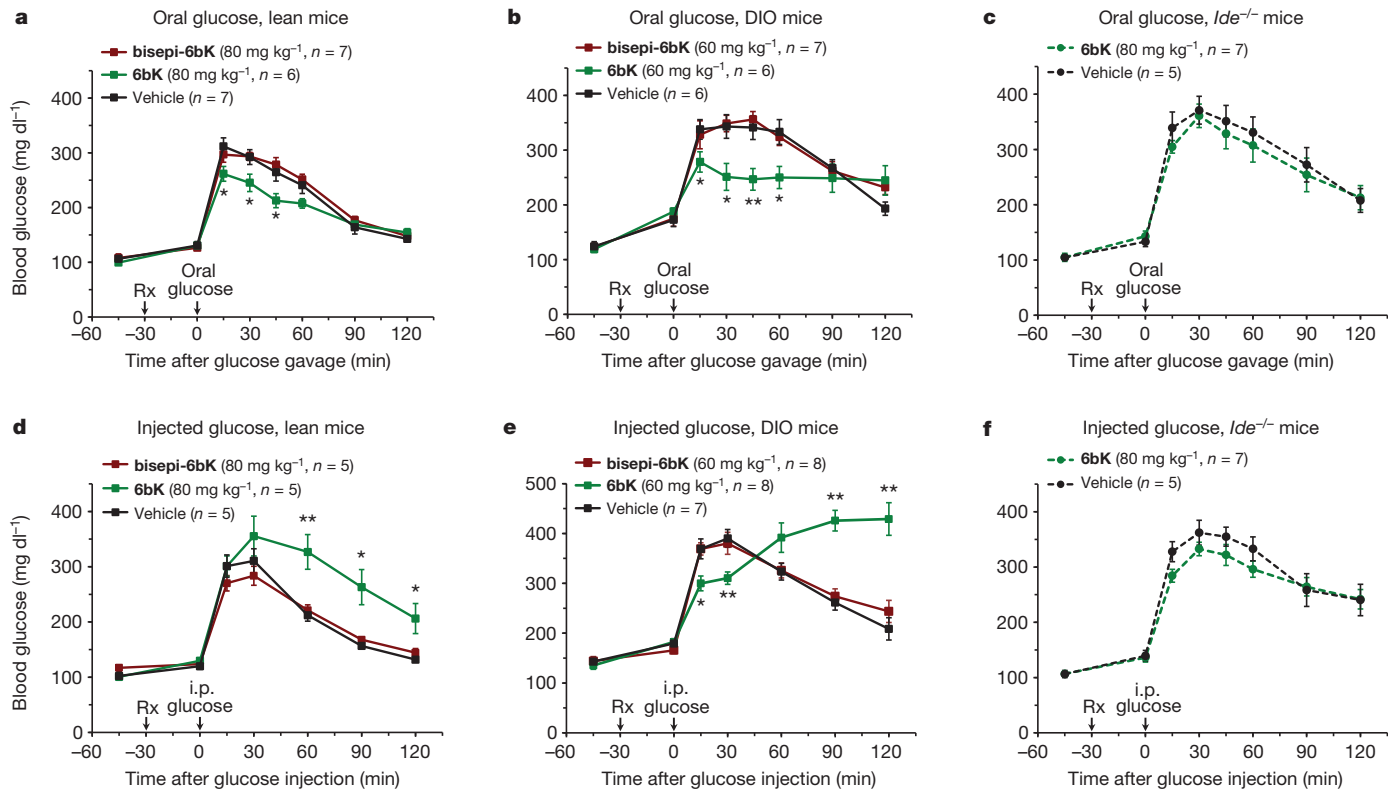


Figure 2 | Physiological consequences of acute IDE inhibition by 6bK on glucose tolerance in lean, DIO and *Ide*^{-/-} mice. **a, b,** Oral glucose tolerance during acute IDE inhibition. **a,** Male C57BL/6J lean (25 g) mice were treated (at time Rx) with a single i.p. injection of IDE inhibitor **6bK**, inactive control **bisepi-6bK**, or vehicle alone, 30 min before glucose gavage (3.0 g kg⁻¹). **b,** DIO mice (35–45 g) were treated with **6bK**, and inactive control **bisepi-6bK** or vehicle alone 30 min before glucose gavage (3.0 g kg⁻¹). **c,** Mice lacking the target (*Ide*^{-/-}) treated with **6bK** followed by oral glucose (3.0 g kg⁻¹) produce a response comparable to that of vehicle-treated

Ide^{-/-} mice. **d, e,** Glucose tolerance phenotypes after i.p. injection of glucose (1.5 g kg⁻¹) in, respectively, lean (**d**) and DIO (**e**) male mice treated with **6bK**, inactive **bisepi-6bK**, or vehicle alone. **f,** Mice lacking IDE treated with **6bK** followed by IPGTT (1.5 g kg⁻¹) produce a response comparable to that of vehicle-treated *Ide*^{-/-} mice. Area under the curve (AUC) calculations are shown in Extended Data Fig. 9. Data: mean ± s.e.m.; **P* < 0.05, ***P* < 0.01 in two-tail Student's *t*-test. Data shown in **a, b, d** and **e** are representative of two or more independent studies. Control studies using knockout mice were performed once.

glucagon secretion^{24,25}. Pramlintide (Symlin) is a synthetic amylin derivative used clinically to control post-prandial glucose levels^{20,21}. To determine the effects of IDE inhibition on endogenous amylin signalling, we measured gastric emptying efficiency, an amylin-mediated process²⁵. Mice treated with **6bK** exhibited twofold slower gastric emptying compared to vehicle and **bisepi-6bK**-treated controls (Fig. 3e). Importantly, co-administration of AC187, a 25-mer peptide amylin receptor antagonist²⁵, blocked the effects of **6bK** on gastric emptying (Fig. 3e). Collectively, these data reveal that IDE inhibition can slow post-prandial gastric emptying by modulating amylin signalling *in vivo* at physiologically relevant levels. Amylin-mediated effects on gastric emptying and satiety have been recognized to have therapeutic relevance^{20,21}, and our results demonstrate a small-molecule-based strategy to modulate amylin signalling (Extended Data Fig. 10).

Higher glucagon levels on **6bK** treatment provide a possible explanation for impaired glucose tolerance during IPGTT experiments. Substrates are processed by IDE at rates dependent on their relative concentrations (Supplementary Discussion). We observed two- to four-fold higher insulin levels during OGTTs than IPGTTs, consistent with the incretin effect^{16–19} (Extended Data Fig. 6). During an OGTT, IDE inhibition therefore results primarily in an increase in insulin signalling and lower blood glucose levels. In contrast, during an IPGTT, insulin secretion levels are lower^{16–19}, IDE processes proportionally more glucagon, and the loss of IDE activity thus results in higher glucagon and glucose levels (Figs 2d, e and 3a; Supplementary Discussion)^{16–19}.

To test this hypothesis, we repeated the glucose tolerance experiments using *Gcgr*^{-/-} mice that lack the G-protein-coupled glucagon receptor (Fig. 3f and g)²⁶. Treatment of *Gcgr*^{-/-} mice with **6bK** followed by an

OGTT resulted in improved glucose tolerance as expected (Fig. 3f), consistent with a model in which insulin signalling in these mice is intact and regulated by IDE. In contrast, glucose tolerance in *Gcgr*^{-/-} mice following IPGTT was not impaired by **6bK** treatment, consistent with a model in which glucagon signalling is responsible for driving higher glucose levels in wild-type mice on IDE inhibition during IPGTTs (compare Figs 3g and 2d). In addition, **6bK** treatment augmented the expression of liver gluconeogenic markers regulated by endogenous glucagon signalling following a pyruvate injection (Supplementary Discussion, Extended Data Fig. 9). These results collectively show that the ability of IDE to regulate glucagon *in vivo* can account for the impaired glucose tolerance observed during IPGTTs (see the Supplementary Discussion on the hyperglycaemic rebound phase in DIO mice).

The discovery and application of the physiologically active IDE inhibitor **6bK** reveals that transient IDE inhibition can improve glucose tolerance under conditions that mimic the intake of a meal. In the context of recent genetics studies identifying *IDE* as a diabetes susceptibility gene^{3,4}, our findings establish the potential of IDE as a target for the treatment of diabetes^{1,2}. Our study also reveals a new requirement for therapeutic strategies that target IDE—namely, that transient IDE inhibition during meals, rather than chronic treatment, is desirable to minimize elevation of glucagon signalling (Extended Data Fig. 10)^{21,27}. Plausible pharmacological strategies to circumvent elevation of post-prandial glucagon levels include: (1) development of fast- and short-acting pre-meal IDE inhibitors²; (2) combination therapy of an IDE inhibitor with glucagon-lowering incretin therapies^{20,21}; and (3) co-administration with glucagon receptor antagonists²⁷. Moreover, existing anti-diabetic therapeutics when co-administered with an IDE inhibitor

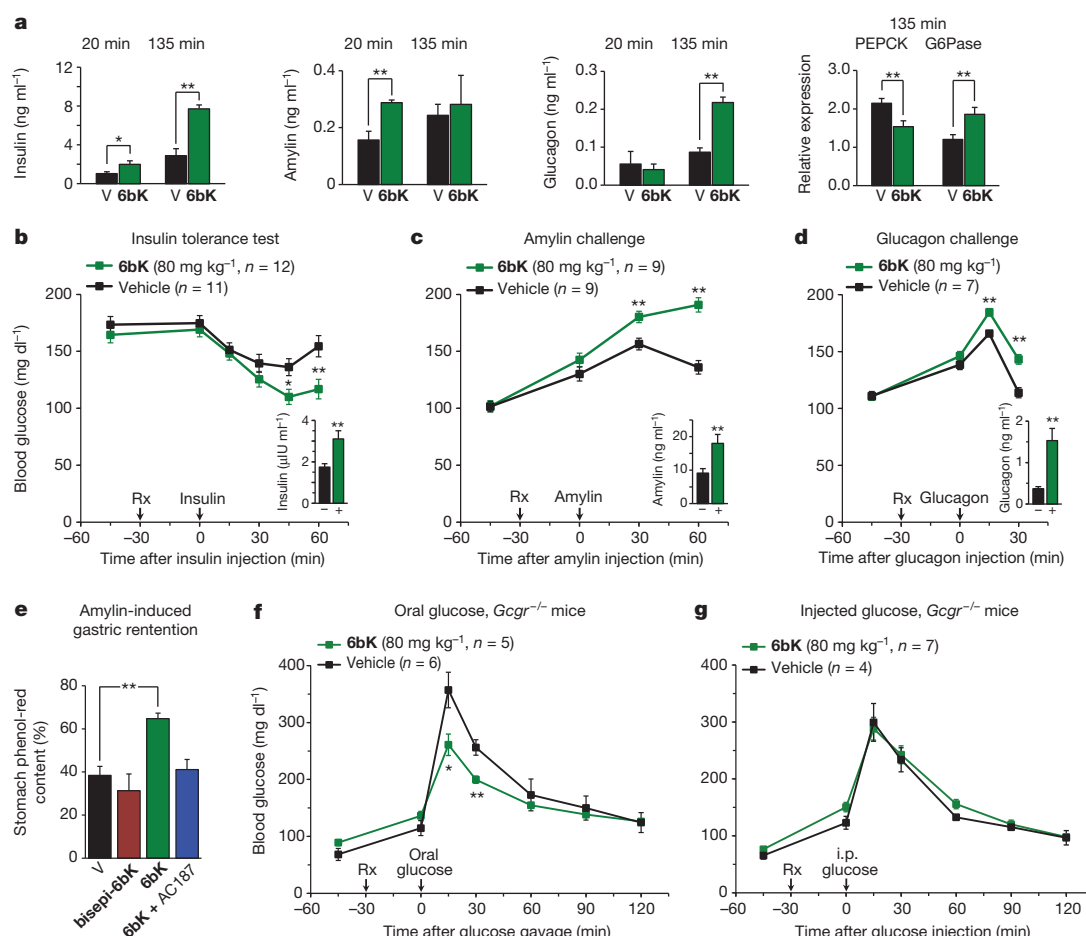


Figure 3 | Acute IDE inhibition affects the abundance of multiple hormone substrates and their corresponding effects on blood glucose levels. **a**, Plasma hormone measurements at 20 min and 135 min after i.p.-injected glucose tolerance tests (IPGTTs) on DIO mice (Fig. 2e) and RT-PCR analysis of DIO liver samples collected at 135 min after IPGTT. RT-PCR reveals 50% higher glucose-6-phosphatase (G6Pase) and 30% lower phosphoenolpyruvate carboxykinase (PEPCK) transcript levels for the **6bK**-treated cohort (**6bK**; $n = 7$) versus vehicle-only controls (V; $n = 7$). **b–d**, Blood glucose responses and injected hormone abundances in lean mice 30 min after treatment with **6bK** or vehicle alone. **b**, Insulin (0.25 U kg^{-1} , subcutaneous) after 5-h fast. **c**, Amylin ($250 \mu\text{g kg}^{-1}$, subcutaneous) after overnight fast. **d**, Glucagon ($100 \mu\text{g kg}^{-1}$, subcutaneous) after overnight fast. Trunk blood was

collected at the last time-points for plasma hormone measurements (insets; IU, international units). **e**, Acute IDE inhibition slows gastric emptying through amylin signalling. Fasted WT mice were given oral glucose supplemented with 0.1 mg ml^{-1} phenol red 30 min after treatment with **6bK** alone ($n = 6$), **6bK** co-administered with the amylin receptor antagonist²⁵ AC187 (3 mg kg^{-1} i.p., $n = 6$), vehicle alone (V; $n = 6$) or inactive **bisepi-6bK** ($n = 4$). **f, g**, G-protein-coupled glucagon receptor knockout mice ($Gcgr^{-/-}$) treated with IDE inhibitor **6bK** display altered glucose tolerance relative to vehicle-treated mice if challenged with oral glucose (3.0 g kg^{-1} ; **f**) but not i.p. injected glucose (1.5 g kg^{-1} ; **g**). Data: mean \pm s.e.m.; * $P < 0.05$, ** $P < 0.01$ in two-tail Student's t -test. Data shown in **a–e** are representative of two or more independent studies. Studies using $Gcgr^{-/-}$ knockout mice were performed once.

may result in synergistic effects²⁸; indeed **6bK** treatment results in stronger improvement in oral glucose tolerance when co-administered with the DPP4 inhibitor sitagliptin (Extended Data Fig. 8)^{19,20}. In addition, the IDE•**6b** structure raises the possibility of developing IDE inhibitors that selectively impede insulin degradation without affecting glucagon degradation. These findings collectively inform and motivate additional studies to progress these discoveries towards new diabetes therapeutics.

METHODS SUMMARY

The *in vitro* selection of the DNA-templated library⁸ used $20 \mu\text{g}$ His₆-tagged mouse IDE immobilized on cobalt magnetic beads (Invitrogen). IDE inhibition was assayed using the fluorogenic peptide Mca-RPPGFSAFK(Dnp)-OH (R&D), confirmed using an anti-insulin antibody time-resolved FRET assay (Cysbio), and a LCMS assay for CGRP cleavage fragments in plasma⁹. Macrocytic inhibitors were synthesized by Fmoc-based solid-phase synthesis and purified by HPLC. LCMS quantitation of **6bK** in biological samples was performed using **6bK** synthesized with $^{13}\text{C}_6$, $^{15}\text{N}_2$ lysine (Sigma-Aldrich).

Wild-type lean and DIO C57BL/6J age-matched male mice (Jackson Laboratories) were used at 14–16, and 24–26 weeks respectively (>20 weeks of high-fat diet). $Gcgr^{-/-}$ and $Ide^{-/-}$ mice were fully backcrossed to the C57BL/6J line, bred from heterozygous mice, and used between 11 and 21 weeks. Animals were fasted overnight

14 h for all experiments, except for the insulin tolerance test, which required 5 h of fasting during the morning. Blood glucose was measured from tail nicks using AccuCheck (Aviva) meters. Trunk blood was obtained for plasma hormone measurements using the Multiplexed Mouse Metabolic Hormone panel (Milliplex, EMD Millipore) on a Luminex FlexMap 3D instrument.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 August 2013; accepted 27 March 2014.

Published online 21 May 2014.

1. Duckworth, W. C., Bennett, R. G. & Hamel, F. G. Insulin degradation: progress and potential. *Endocr. Rev.* **19**, 608–624 (1998).
2. Mirsky, I. A. & Broh-Kahn, R. H. The inactivation of insulin by tissue extracts; the distribution and properties of insulin inactivating extracts. *Arch. Biochem.* **20**, 1–9 (1949).
3. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
4. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
5. Farris, W. *et al.* Insulin-degrading enzyme regulates the levels of insulin, amyloid beta-protein, and the beta-amyloid precursor protein intracellular domain *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 4162–4167 (2003).

6. Abdul-Hay, S. O. *et al.* Deletion of insulin-degrading enzyme elicits antipodal, age-dependent effects on glucose and insulin tolerance. *PLoS ONE* **6**, e20818 (2011).
7. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305**, 1601–1605 (2004).
8. Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for *in vitro* selection. *J. Am. Chem. Soc.* **130**, 15611–15626 (2008).
9. Kim, Y. G., Lone, A. M., Nolte, W. M. & Saghatelian, A. Peptidomics approach to elucidate the proteolytic regulation of bioactive peptides. *Proc. Natl Acad. Sci. USA* **109**, 8523–8527 (2012).
10. Leissring, M. A. *et al.* Designed inhibitors of insulin-degrading enzyme regulate the catabolism and activity of insulin. *PLoS ONE* **5**, e10504 (2010).
11. Malito, E. *et al.* Molecular bases for the recognition of short peptide substrates and cysteine-directed modifications of human insulin-degrading enzyme. *Biochemistry* **47**, 12822–12834 (2008).
12. Malito, E., Hulise, R. E. & Tang, W. J. Amyloid beta-degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin. *Cell. Mol. Life Sci.* **65**, 2574–2585 (2008).
13. Shen, Y., Joachimiak, A., Rosner, M. R. & Tang, W. J. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **443**, 870–874 (2006).
14. Manolopoulou, M., Guo, Q., Malito, E., Schilling, A. B. & Tang, W. J. Molecular basis of catalytic chamber-assisted unfolding and cleavage of human insulin by human insulin-degrading enzyme. *J. Biol. Chem.* **284**, 14177–14188 (2009).
15. Stella, V. J. & He, Q. Cyclodextrins. *Toxicol. Pathol.* **36**, 30–42 (2008).
16. Andrikopoulos, S., Blair, A. R., Deluca, N., Fam, B. C. & Proietto, J. Evaluating the glucose tolerance test in mice. *Am. J. Physiol. Endocrinol. Metab.* **295**, E1323–E1332 (2008).
17. Ahrén, B., Winzell, M. S. & Pacini, G. The augmenting effect on insulin secretion by oral versus intravenous glucose is exaggerated by high-fat diet in mice. *J. Endocrinol.* **197**, 181–187 (2008).
18. Winzell, M. S. & Ahren, B. The high-fat diet-fed mouse: a model for studying mechanisms and treatment of impaired glucose tolerance and type 2 diabetes. *Diabetes* **53** (Suppl 3), S215–S219 (2004).
19. Drucker, D. J. The biology of incretin hormones. *Cell Metab.* **3**, 153–165 (2006).
20. Riddle, M. C. & Drucker, D. J. Emerging therapies mimicking the effects of amylin and glucagon-like peptide 1. *Diabetes Care* **29**, 435–449 (2006).
21. Mooradian, A. D. & Thurman, J. E. Drug therapy of postprandial hyperglycaemia. *Drugs* **57**, 19–29 (1999).
22. Duckworth, W. C. & Kitabchi, A. E. Insulin and glucagon degradation by the same enzyme. *Diabetes* **23**, 536–543 (1974).
23. Bennett, R. G., Duckworth, W. C. & Hamel, F. G. Degradation of amylin by insulin-degrading enzyme. *J. Biol. Chem.* **275**, 36621–36625 (2000).
24. Young, A. Effects on plasma glucose and lactate. *Adv. Pharmacol.* **52**, 193–208 (2005).
25. Gedulin, B. R., Jodka, C. M., Herrmann, K. & Young, A. A. Role of endogenous amylin in glucagon secretion and gastric emptying in rats demonstrated with the selective antagonist, AC187. *Regul. Pept.* **137**, 121–127 (2006).
26. Gelling, R. W. *et al.* Lower blood glucose, hyperglucagonemia, and pancreatic alpha cell hyperplasia in glucagon receptor knockout mice. *Proc. Natl Acad. Sci. USA* **100**, 1438–1443 (2003).
27. Unger, R. H. & Cherrington, A. D. Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. *J. Clin. Invest.* **122**, 4–12 (2012).
28. Sadry, S. A. & Drucker, D. J. Emerging combinatorial hormone therapies for the treatment of obesity and T2DM. *Nature Rev. Endocrinol.* **9**, 425–433 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by NIH/NIGMS (R01 GM065865 (D.R.L.), R00 GM080097 (M.A.S.), R01 GM81539 (W.-J.T.), T32 GM008444 (Z.H.F.), F30 CA174152 (Z.H.F.), DP2 OD002374 (A.S.)), Howard Hughes Medical Institute (D.R.L.), Diabetes and Cancer Centers of Albert Einstein College of Medicine (M.J.C.), American Diabetes Association no. 7-11-CD-06 (M.A.L.), Burroughs Wellcome Fund CABS (A.S.), and the Searle Scholars Program (A.S.). The Fonds de Recherche en Santé du Québec (FRSQ) and the Alfred Bader Fund supported J.P.M. We thank C. Russ and H. Spurling (Broad Institute) and C. Daly (FAS Center for Systems Biology) for DNA sequencing assistance. We are grateful to S. Johnston and C. Mosher (Broad Institute) for **6bK** stability measurements. W. Nolte provided mouse IDE, L. McCord purified CF-IDE and Y.-G. Kim performed CGRP cleavage assays. We are grateful to A. Badran, E. Homan, A. M. Lone and M. Leidl (Harvard University) for experimental assistance. We thank B. Kahn and N. Gray for discussions.

Author Contributions J.P.M., A.M., Z.H.F., R.E.K. and X.Q.D. performed the experiments. M.A.L. provided the *Ide*^{-/-} mice. W.-J.T. provided IDE protein for structural studies. M.J.C. supervised the *Gcgr*^{-/-} studies. M.A.S. supervised the IDE-**6b** structural studies. A.S. supervised the pharmacological validation of **6bK** and the *in vivo* studies. D.R.L. supervised the discovery of IDE inhibitors, the pharmacological studies and the *in vivo* studies. All authors analysed the data and wrote the manuscript.

Author Information The coordinates and the structure factors of the IDE-**6b** complex have been deposited in the Protein Data Bank under the accession code 4LTE. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R.L. (drliu@fas.harvard.edu) and A.S. (saghatelian@chemistry.harvard.edu).

Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk

Katrin D. Mayer-Barber¹, Bruno B. Andrade¹, Sandra D. Oland¹, Eduardo P. Amaral^{1,2}, Daniel L. Barber³, Jacqueline Gonzales⁴, Steven C. Derrick⁵, Rui Shi⁶, Nathella Pavan Kumar^{7,8}, Wang Wei⁶, Xing Yuan⁶, Guolong Zhang⁹, Ying Cai⁴, Subash Babu^{7,10}, Marta Catalfamo¹¹, Andres M. Salazar¹², Laura E. Via⁴, Clifton E. Barry III⁴ & Alan Sher¹

Tuberculosis remains second only to HIV/AIDS as the leading cause of mortality worldwide due to a single infectious agent¹. Despite chemotherapy, the global tuberculosis epidemic has intensified because of HIV co-infection, the lack of an effective vaccine and the emergence of multi-drug-resistant bacteria^{2–5}. Alternative host-directed strategies could be exploited to improve treatment efficacy and outcome, contain drug-resistant strains and reduce disease severity and mortality⁶. The innate inflammatory response elicited by *Mycobacterium tuberculosis* (Mtb) represents a logical host target⁷. Here we demonstrate that interleukin-1 (IL-1) confers host resistance through the induction of eicosanoids that limit excessive type I interferon (IFN) production and foster bacterial containment. We further show that, in infected mice and patients, reduced IL-1 responses and/or excessive type I IFN induction are linked to an eicosanoid imbalance associated with disease exacerbation. Host-directed immunotherapy with clinically approved drugs that augment prostaglandin E2 levels in these settings prevented acute mortality of Mtb-infected mice. Thus, IL-1 and type I IFNs represent two major counter-regulatory classes of inflammatory cytokines that control the outcome of Mtb infection and are functionally linked via eicosanoids. Our findings establish proof of concept for host-directed treatment strategies that manipulate the host eicosanoid network and represent feasible alternatives to conventional chemotherapy.

The mechanisms by which IL-1-driven inflammation promotes host resistance against Mtb *in vivo* have not been elucidated and offer novel targets for immunological intervention in tuberculosis (TB). Despite the high susceptibility of IL-1-deficient mice to Mtb infection, established mediators of host resistance, such as interferon (IFN)- γ , tumour necrosis factor (TNF)- α , inducible nitric oxide synthase (iNOS) or IL-12p40 were not diminished in the absence of IL1R1 when assessed *in vitro*, *in vivo* at the single-cell level or in bronchoalveolar lavage fluid (BALF) (Extended Data Fig. 1a–c)^{8,9}. Instead, we observed that increased bacterial loads in BALF were accompanied by an eicosanoid imbalance (Fig. 1a, b, Extended Data Fig. 1d). Eicosanoids are lipid mediators derived from arachidonic acid that include prostaglandins, resolvins, lipoxins and leukotrienes¹⁰. Cyclooxygenase-2 (COX-2, gene also known as *Ptgs2*) and 5-lipoxygenase (5-LO, gene also known as *Alox5*) or 12/15-lipoxygenase (12/15-LO) compete for arachidonic acid to generate COX-dependent prostaglandins or LO-dependent lipoxins and leukotrienes. *Alox5*^{-/-} mice have been reported to be more resistant and prostaglandin E synthase (*Ptges*)-deficient mice more susceptible to Mtb infection^{11–13}. In the absence of IL-1 signalling, we found that prostaglandin E2 (PGE2) and prostaglandin F2 α (PGF2 α) were significantly reduced in BALF of Mtb-infected *Il1r1*^{-/-} mice, whereas the

5-LO products LXA4 and LTB4 were increased compared to wild-type animals (Fig. 1b, Extended Data Fig. 1d).

We then purified myeloid cells from infected wild-type lungs by fluorescence-activated cell sorting (FACS), and found that they produced large amounts of PGE2 *ex vivo* (Extended Data Fig. 1e). We next used mixed bone marrow chimaeras to analyse the requirement for IL1R1 on pulmonary CD68^{pos} myeloid cells for COX-2 production *in vivo*. *Il1r1*^{-/-} cells expressed significantly less COX-2 protein, indicating that IL-1 acts directly on myeloid cells for optimal COX-2 induction (Extended Data Fig. 1b, f) independently of pulmonary bacterial loads. *In vitro*, Mtb-infected *Il1r1*^{-/-} bone-marrow-derived macrophages (BMDM) failed to control bacterial growth compared to wild-type BMDM (Fig. 1c) and the increased colony forming units in *Il1r1*^{-/-} BMDM or bone-marrow dendritic cell cultures were associated with decreased PGE2 (Extended Data Fig. 2a, b). In addition, we observed a significant increase in both extracellular bacterial as well as infected macrophage numbers in *Il1a*, *Il1b*^{-/-} cultures (Extended Data Fig. 2c, d). Add-back of exogenous IL-1 α , IL-1 β or PGE2 reduced extracellular bacteria to wild-type levels (Fig. 1d, Extended Data Fig. 2e–g) and restored control of bacterial growth in IL-1-deficient BMDM (Fig. 1e and Extended Data Fig. 3a).

IL-1-dependent bacterial control was lost when COX-2 was inhibited with valdecoxib and restored upon PGE2 addition (Fig. 1e and Extended Data Fig. 3b). Of note, valdecoxib increased bacterial loads and inhibited PGE2 synthesis in infected wild-type BMDM as was observed in BMDM derived from B6 mice deficient in the enzymatic activity of COX-2 (*Ptgs2*(Y385F) mice) (Extended Data Fig. 3c). *In vivo*, absence of COX-2 enzymatic activity did not affect pulmonary and serum levels of PGE2 or lung bacterial loads in *Ptgs2*(Y385F) mice at 4 weeks after Mtb infection, consistent with a compensatory role for COX-1, which can generate PGE2 when COX-2 is absent¹⁴ (Fig. 1f, Extended Data Fig. 3d, e). Nonetheless, at later stages of Mtb infection, *Ptgs2*(Y385F) mice exhibited increased susceptibility compared to wild-type animals (Fig. 1g). We next examined whether IL-1 also drives PGE2 synthesis in infected primary human monocyte-derived macrophages (MDM). Blockade of IL1R1 signalling resulted in reduced PGE2 levels, whereas addition of exogenous IL-1 α or IL-1 β further increased PGE2 synthesis during Mtb infection (Fig. 1h). Although additional PGE2-independent mechanisms of IL-1-mediated protection are likely to exist, our data demonstrate *in vitro* and *in vivo* that IL-1 α and IL-1 β , acting partially through COX-2, trigger PGE2 synthesis, which in turn regulates the containment and replication of bacilli during Mtb infection.

In many settings type I IFNs promote bacterial virulence and disease exacerbation^{15,16}. Hypervirulence of certain Mtb strains correlates with

¹Immunobiology Section, Laboratory of Parasitic Diseases (LPD), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, Maryland 20892, USA.

²Department of Immunology, Biomedical Sciences Institutes, University of Sao Paulo, 05508-900 Sao Paulo, Brazil. ³T Lymphocyte Biology Unit, LPD, NIAID, NIH, Bethesda, Maryland 20892, USA.

⁴Tuberculosis Research Section, Laboratory of Clinical Infectious Disease, NIAID, NIH, Bethesda, Maryland 20892, USA. ⁵Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892, USA. ⁶Henan Chest Hospital, 450003 Zhengzhou, China. ⁷NIH, International Center for Excellence in Research, 600 031 Chennai, India. ⁸National Institute for Research in Tuberculosis (NIRT), 600 031 Chennai, India. ⁹Sino-US International Research Center for Tuberculosis, and Henan Public Health Center, 450003 Zhengzhou, China. ¹⁰Helminth Immunology Section, LPD, NIAID, NIH, Bethesda, Maryland 20892, USA. ¹¹Clinical and Molecular Retrovirology Section, Laboratory of Immunoregulation, NIAID, NIH, Bethesda, Maryland 20892, USA. ¹²Oncovir Inc., Washington, Washington DC 20008, USA.

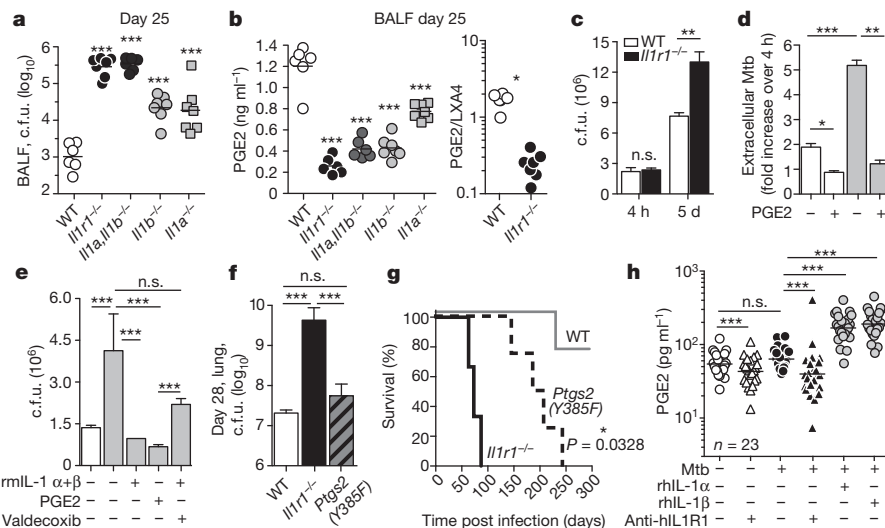


Figure 1 | IL-1 triggers PGE2 synthesis during Mtb infection. **a**, Bacterial loads (colony forming units, c.f.u.); **b**, PGE2 concentrations and PGE2/LXA ratio 25 days after aerosol exposure to Mtb in BALF of C57BL/6 wild type (WT), *Il1r1*^{-/-}, *Il1α, Il1β*^{-/-}, *Il1β*^{-/-} or *Il1α*^{-/-} mice ($n = 4-8$). **c**, c.f.u. in Mtb-infected (m.o.i. = 3) WT or *Il1r1*^{-/-} BMDM cultures. **d**, Mtb-RFP-infected (m.o.i. = 3) WT and *Il1α, Il1β*^{-/-} BMDMs cultured with or without PGE2. Extracellular bacteria were analysed by FACS (day 5). **e**, c.f.u. after *in vitro* infection (m.o.i. = 3) of WT or *Il1α, Il1β*^{-/-} BMDM in the presence or absence of IL-1, PGE2 or the COX-2 inhibitor valdecoxib. rmIL-1α, recombinant murine IL-1α. **f**, Lung c.f.u. of 4-week Mtb-infected WT, *Il1r1*^{-/-} or *Ptgs2*(Y385F)

mice ($n = 5$). **a-f**, Data are representative of two independent experiments. * $P \leq 0.05$; ** $P < 0.005$; *** $P < 0.0005$; compared as indicated in figure by lines or to WT control groups. Error bars denote s.d. (Mann-Whitney test). **g**, Survival of WT, *Il1r1*^{-/-} or *Ptgs2*(Y385F) animals after aerosol challenge with 25–40 c.f.u. Data shown are from one experiment ($n = 4$, Mantel-Cox test). **h**, PGE2 concentration in supernatants of MDM from 23 healthy donors, 24 h after Mtb infection in the presence or absence of rhIL-1α, rhIL-1β or anti-hIL1R1 neutralizing antibody. rhIL-1α, recombinant human IL-1α; anti-hIL1R1, anti-human IL-1R1. Differences were compared as indicated by lines (Wilcoxon matched pairs test).

enhanced necrosis, type I IFN production and Mtb-infected *Ifnar1*^{-/-} mice harbour fewer bacteria compared to wild-type animals^{8,17–19}. Type I IFNs subvert anti-tuberculous host defences by inhibiting iNOS,

IL-12p40, IL-1α and IL-1β production, while inducing the immunosuppressive mediators IL-10 and IL-1R antagonist (IL-1Ra, gene also known as *Il1rn*)^{8,19}. Consistent with our previous findings that type I IFNs

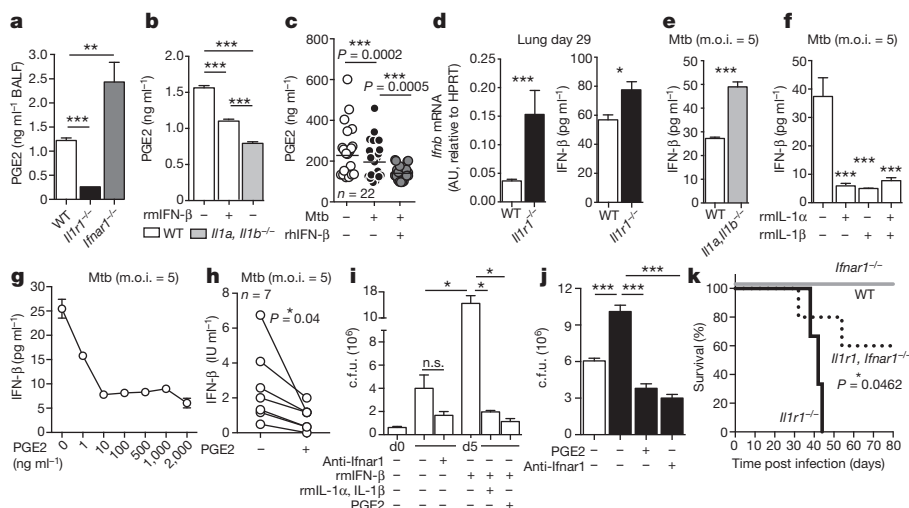


Figure 2 | IL-1 and PGE2 negatively regulate type I IFNs. **a**, **b**, PGE2 measured in BALF of indicated mouse strains 4 weeks post infection (p.i.) ($n = 5$) (**a**) and WT or *Il1α, Il1β*^{-/-} BMDM infected for 40 h (m.o.i. = 5) or in the presence or absence of recombinant murine IFN-β (rmIFN-β; **b**). **c**, PGE2 in supernatants of MDM from 22 healthy donors, 24 h after Mtb infection (m.o.i. = 5) in the presence or absence of rhIFN-β (Wilcoxon matched pairs test). **d**, *Ifnb* mRNA expression and IFN-β protein in lungs of Mtb-infected WT or *Il1r1*^{-/-} mice at indicated time points. AU, arbitrary units ($n = 9$). **e**, **f**, IFN-β protein in supernatants of WT or *Il1α, Il1β*^{-/-} BMDM infected for 40 h (**e**) or WT BMDM in the presence or absence of rmIL-1α or IL-1β (**f**). **g**, IFN-β protein concentration in supernatants WT BMDM (40 h p.i. Mtb) treated with increasing concentrations of PGE2. **h**, IFN-β protein in supernatants of

MDM from 7 healthy donors, 24 h after Mtb infection with or without PGE2 (Wilcoxon matched pairs test). **i**, c.f.u. at indicated time points of Mtb-infected (m.o.i. = 5) WT BMDM cultures treated with anti-IFNAR1 monoclonal antibody, rmIL-1α or IL-1β, rmIFN-β or PGE2. **j**, c.f.u. after 5 days of Mtb infection (m.o.i. = 5) of *Il1r1*^{-/-} BMDM treated with anti-mIFNAR1 mAb or PGE2. **k**, Survival of WT, *Il1r1*^{-/-}, *Ifnar1*^{-/-} or *Il1r1, Ifnar1*^{-/-} double-deficient animals after Mtb infection ($n = 5$, Mantel-Cox test). Data shown are representative of two (**a**, **b**, **d**, **e**, **i**, **k**) or three (**j**) independent experiments. * $P \leq 0.05$; ** $P < 0.005$; *** $P < 0.0005$; compared as indicated in figure by lines or to WT control groups. Error bars denote s.d. (Mann-Whitney test).

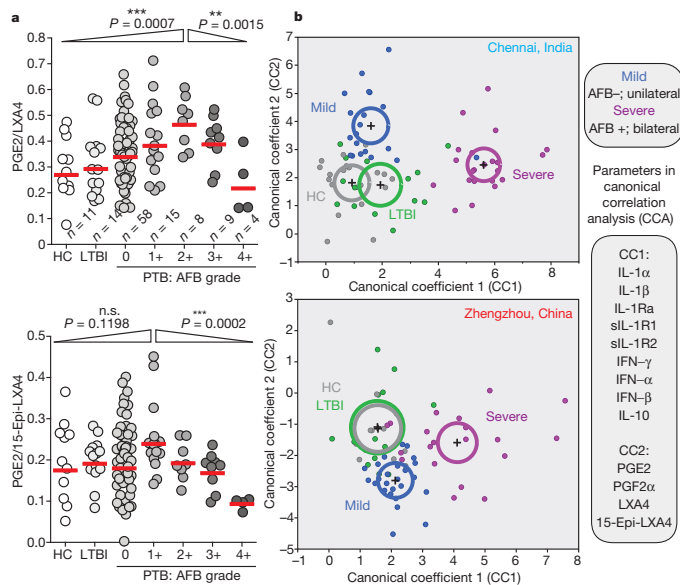


Figure 3 | IL-1, interferon and eicosanoid pathways are engaged in active TB disease. **a**, Mann-Kendall linear trend post hoc comparison of ratios of PGE2/LXA4 and PGE2/15-Epi-LXA4 according to AFB grade in the Chinese cohort. Each dot represents one individual with triangles depicting directionality of trend. **b**, PTB patients were stratified into mild (AFB-, unilateral lung lesions) and severe disease (AFB ≥ 1+, and bilateral lung lesions) categories and canonical correlation analysis (CCA) was performed independently for Indian and Chinese cohorts. Indicated parameters were used to differentiate PTB from HC and LTBI groups as well as clinical severity within PTB group based on combined canonical coefficients. sIL-1R, soluble IL-1R.

antagonize the IL-1 pathway during Mtb infection⁸, absence of type I IFN signalling resulted in increased PGE2 and IL-1β and decreased IL1Ra in BALF (Fig. 2a and Extended Data Fig. 4a). Mtb-infected wild-type BMDM produced significantly less PGE2 when exogenous IFN-β was present (Fig. 2b). Addition of IFN-β to infected human MDM also significantly reduced PGE2 (Fig. 2c, Extended Data Fig. 4b), yet supplementation of IFN-β-treated cultures with IL-1 failed to restore PGE2 levels (Extended Data Fig. 4c). Together these observations suggested that, in addition to IL-1 itself and even in the presence of exogenous IL-1, type I IFNs potentially counter-regulate the prostaglandin axis in Mtb-infected human and murine cells.

We next asked whether in the opposing direction IL-1 and PGE2 inhibit the type I IFN pathway, which could represent an important mechanism of IL-1/PGE2-dependent bacterial control. Indeed, in Mtb-infected *Il1r1*^{-/-} mice pulmonary messenger RNA and protein levels of IFN-β were significantly elevated compared to wild-type mice (Fig. 2d). Mtb-infected *Il1a*, *Il1b*^{-/-} BMDM also expressed significantly elevated *Ifna*, *Ifnb*, IFN-inducible *Il1rn* (IL1Ra) and *Il1r2* decoy receptor mRNA as well as IFN-β protein (Extended Data Fig. 4d and Fig. 2e). Exogenous IL-1α and IL-1β inhibited IFN-β protein in wild-type macrophages and reduced mRNA expression in *Il1a*, *Il1b*^{-/-} BMDM (Fig. 2f, Extended Data Fig. 4e). Importantly, exogenous PGE2 also suppressed type I IFN production in Mtb-infected murine and human macrophages (Fig. 2g, h). Addition of either IL-1 or PGE2 to IFN-β-treated BMDM cultures reversed IFN-β-dependent loss of bacterial control (Fig. 2i). IL-1 failed to reverse PGE2 inhibition by type I IFNs, yet fully restored bacterial control, suggesting the existence of other IL-1-driven anti-mycobacterial mechanisms in addition to PGE2 regulation.

To investigate if the IL-1-type I IFN counter-regulatory pathway has an important function in IL-1-dependent bacterial control, we neutralized type I IFN in infected *Il1r1*^{-/-} BMDM. Blocking type I IFN signalling significantly reduced bacterial growth in *Il1r1*^{-/-} macrophages (Fig. 2j). Importantly, mice doubly deficient in both IL1R1 and IFNAR1

(*Il1r1*, *Ifnar1*^{-/-}) were significantly less susceptible to Mtb infection than *Il1r1*^{-/-} single-deficient animals (Fig. 2k). These findings reveal suppression of type I IFNs and their pro-bacterial activity as a major mechanism of IL-1- and PGE2-mediated host resistance against Mtb.

In previous studies both the development of active TB in latently infected individuals and the extent of radiographic disease among active TB patients were found to be closely associated with a type I IFN gene signature^{20–22}. Therefore we speculated that the IL-1-type I IFN balance might be particularly relevant in active TB patients with more severe disease presentations, than in latent or asymptomatic cases²³. We retrospectively measured components of the IL-1, eicosanoid and IFN pathways in plasma of patients from Chennai, India, stratified into healthy controls (HC), latently infected individuals (LTBI) and pulmonary TB (PTB) cases based on their acid fast (AFB) sputum classification (Supplementary Tables 1, 2, 6). With increasing disease status, IL-1α was reduced whereas PGE2, LXA4 and 15-Epi-LXA4 showed a significant trend to increase (Supplementary Table 6). We next measured the same parameters in a cohort from Zhengzhou, China, and found similar trends for IL-1α, PGE2 and 15-Epi-LXA4 (Supplementary Tables 3, 5, 7). The decreases in plasma IL-1α observed here diverge from previous findings in BALF or alveolar macrophages of active TB patients, where IL-1β was shown to be elevated^{24–27}. This may reflect differences in systemic versus pulmonary cytokine levels or between the IL-1 species because we observed a similar dichotomy in mice infected with Mtb (Extended Data Fig. 5). The finding of increased PGE2 concentrations in PTB compared to HC and/or LTBI patients was unanticipated based on our murine data. Nevertheless, when we investigated the eicosanoid balance within the PTB group, as proposed above, we found a significant linear trend for PGE2/LXA4 or PGE2/15-Epi-LXA4 ratios to be decreased with higher sputum grades (Fig. 3a). We then further stratified the PTB group into patients with mild or severe disease (Fig. 3b, Supplementary Table 5). Importantly, in both study sites the IL-1, eicosanoid and IFN pathways faithfully distinguished PTB patients based on disease severity after canonical correlation analysis (CCA)²⁸ (Fig. 3b, Supplementary Table 8). Thus, clinical data from two independent study cohorts support the hypothesis that TB severity is associated with changes in IL-1, type I IFNs and eicosanoids thus supporting a link between these pathways in active tuberculosis.

On the basis of these observations we proposed that a host-directed therapy (HDT) targeting eicosanoids would be most effective in active disease associated with necrotic lung pathology where high type I IFN conditions dominate. Indeed, when *Il1a*, *Il1b*^{-/-} animals were administered PGE2 together with zileuton, a 5-LO inhibitor (Extended Data Fig. 6a), pulmonary bacterial loads and pathology were reduced and survival enhanced (Fig. 4a, Extended Data Fig. 6b–e). Of note, when the same treatment was given to wild-type animals for up to 6 months bacterial control was not enhanced nor were negative side effects observed (Fig. 4a, Extended Data Fig. 6b, e and data not shown). Importantly, PGE2 and zileuton treatment had no effect on survival of susceptible *Ifng*^{-/-} or *Tnfa*^{-/-} animals (Extended Data Fig. 6f). Thus, the observed treatment is specific for IL-1-mediated host resistance, rather than owing to indirect effects on pathways generally shared by susceptible mice. We next administered either PGE2 or zileuton alone and found that add-back of PGE2 extended survival of *Il1r1*^{-/-} animals, whereas administration of zileuton failed to do so (Extended Data Fig. 6g). Consistent with the latter finding, infected *Il1r1*, *Alox5*^{-/-} mice died with similar kinetics as *Il1r1*^{-/-} animals (Extended Data Fig. 6g). These data indicate that the susceptibility of IL-1-deficient mice is linked to decreased PGE2, rather than to an increase in 5-LO products, and that yet unidentified functions of IL-1 contribute to host resistance against Mtb, because rescue of IL-1-deficient animals, either by PGE2 add-back or type I IFN neutralization, was incomplete.

We then tested the eicosanoid-based HDT in a model using IL-1-replete animals where induction of high type I IFN levels by intranasal administration of polyinosinic-polycytidylic acid stabilized with poly-L-lysine (pICLC) causes uncontrolled disease leading to mortality and

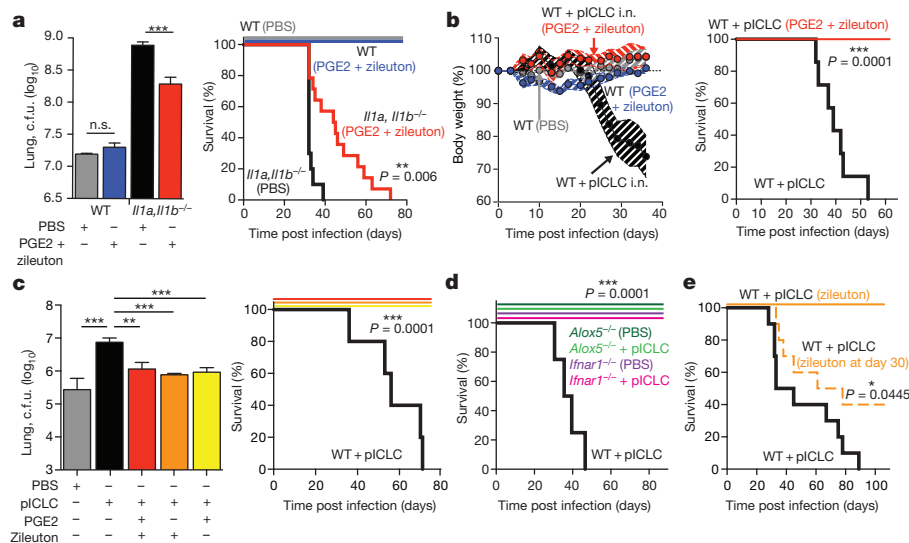


Figure 4 | HDT targeting eicosanoids confers protection in highly susceptible mice associated with high type I IFN responses. **a**, Lung c.f.u. (day 21 p.i.) (left, $n = 10$, error bars denote s.d., Mann–Whitney test) and survival (right, $n = 13$, Mantel–Cox test) between treated and untreated *Il1 α , Il1 β* ^{-/-} mice after aerosol exposure to Mtb of WT or *Il1 α , Il1 β* ^{-/-} treated with or without intranasal (i.n.) PGE2 (twice a week) and zileuton (in drinking water) starting day 1 p.i. Data shown are combined from two independent experiments. **b–e**, Mice were treated i.n. with pICLC or without (PBS) twice a week, starting day 1 after aerosol challenge with Mtb. **b**, Weight loss (left, travelling error bars indicate s.d., $n = 7$) and survival (right) during Mtb infection of WT (pICLC or PBS i.n.) mice treated with or without i.n. PGE2 and zileuton. **c**, Lung c.f.u. 4 weeks p.i. of WT mice treated as indicated (left, $n = 5$, error bars denote s.d. (Mann–Whitney test). Survival (right) of

pICLC-treated WT mice (black) compared to those who also received PGE2 (yellow), zileuton (orange) or both (red). Differences were compared to pICLC group (Mantel–Cox test). Data are representative of three independent experiments ($n = 5$). **d**, Survival of pICLC-treated WT mice (black), *Ifnar1*^{-/-} animals (purple, without pICLC in pink) or *Alox5*^{-/-} mice (dark green, without pICLC in light green). Statistical significance was compared to pICLC group. Data are representative of two independent experiments ($n = 4–6$). **e**, Survival of pICLC-administered WT mice treated with zileuton in drinking water starting day 1 (solid orange line) compared to d30 (dotted orange line). Data shown are combined from two independent experiments ($n = 10$). Statistical significance was compared to pICLC (black) group with Mantel–Cox test. * $P \leq 0.05$; ** $P < 0.005$; *** $P < 0.0005$.

necrotic lung pathology¹⁵. While triggering loss of bacterial control, increased cell death, IFN- α , IFN- β , IL-10 and IL1Ra expression in mouse lungs, pICLC treatment failed to overtly suppress IL-1 or PGE2 (Extended Data Fig. 5), perhaps reflecting an indirect effect of the increased bacterial burden and/or type I IFN induction of IL1Ra. When PGE2 and zileuton were administered together, pICLC-treated mice were fully protected against weight loss and mortality, displaying lower colony forming units as well as decreased necrotic lung pathology (Fig. 4b, Extended Data Fig. 7a). PGE2 or zileuton alone reduced pulmonary colony forming units, which resulted in complete protection against both weight loss and mortality (Fig. 4c, Extended Data Fig. 7b). Importantly, protection directly correlated with increased PGE2 and decreased IFN- β and IL1Ra production, even when animals were given 5-LO blockade alone (Extended Data Fig. 7c, e).

We confirmed that zileuton and PGE2 did not directly inhibit growth of Mtb *in vitro* at concentrations as high as 75 $\mu\text{g ml}^{-1}$ and 50 $\mu\text{g ml}^{-1}$, respectively (data not shown), and co-administration did not interfere with standard antibiotic therapy (Extended Data Fig. 8). Mtb-infected *Ifnar1*^{-/-} and *Alox5*^{-/-} animals were resistant to pICLC-driven disease, confirming that pICLC-driven susceptibility is type I IFN-mediated and that zileuton protects by targeting 5-LO rather than through off-target effects (Fig. 4d, Extended Data Fig. 9a, b). IL-10 is induced by type I IFNs and contributes to its antagonism of IL-1 during Mtb^{8,19}, and accordingly *Il10*^{-/-} animals survived pICLC administration (Extended Data Fig. 9c). LTB4, although 5-LO-dependent and previously implicated in pathology of zebrafish infected with *Mycobacterium marinum*^{29,30}, did not mediate pICLC-driven disease, because *Ltb4r1*^{-/-} mice succumbed with similar kinetics as wild-type controls (Extended Data Fig. 9d). Likewise, 12/15-LO^{-/-} animals displayed the same mortality as wild-type animals, arguing against a role of 12/15-LO (Extended Data Fig. 9d). Because PGE2 supplementation, either by direct addition or 5-LO inhibition with zileuton, completely prevented type I IFN-driven

mortality, it is possible that the HDT protects via PGE2-mediated bacterial control and/or inhibition of type I IFNs rather than suppression of 5-LO products.

Finally, we tested the therapeutic efficacy of 5-LO blockade by initiating Zileuton treatment at 30 days post-infection, when pICLC-treated animals already display 10–20% weight loss (Fig. 4e and data not shown). This resulted in enhanced survival, with 40% of the animals remaining alive at the end of the study. Thus, manipulation of the eicosanoid balance towards PGE2 can both prevent and therapeutically ameliorate disease exacerbation associated with type I IFN expression (Extended Data Fig. 10). Although eicosanoids have been suggested previously as targets for therapeutic intervention in tuberculosis, the data presented here provide *in vivo* proof of concept for this approach in a mammalian model of pulmonary Mtb infection.

METHODS SUMMARY

Eight-to-twelve-week-old C57BL/6 or gene-deficient animals were aerosol infected with Mtb strain H37Rv (100–150 colony forming units per mouse unless stated otherwise). *In vitro* infections of MDM or BMDM were performed at indicated multiplicity of infection (m.o.i.). Eicosanoids and cytokines were measured between 24 and 48 h after *in vitro*, *in vivo* at indicated time points or in plasma of TB patients using commercial enzyme immunoassay (EIA) or enzyme-linked immunosorbent assay (ELISA) kits. For *in vitro* colony forming units determination, supernatants were collected day 5 post infection and cells lysed in distilled H₂O for 5–10 min. Supernatants and cell lysate were then combined to evaluate total mycobacterial growth. Extracellular colony forming units were determined by microscopy or flow cytometry as described in Methods. Antibodies, PCR primers and other pharmacological reagents as well as staining, RT-PCR and cell differentiation protocols are listed in Methods. C57BL/6 mice were administered pICLC intranasally twice a week (6 μg in 30 μl). For HDT mice were given PGE2 intranasally twice a week (2 μg in 30 μl) or zileuton in drinking water *ad libitum* (6 mg ml⁻¹) or both starting 1 day after infection, unless stated otherwise. Patient cohort descriptions and immune mediator measurements are provided in Supplementary Information and Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 September 2013; accepted 16 May 2014.

Published online 25 June 2014.

- WHO. Tuberculosis Fact sheet N°104. <http://www.who.int/mediacentre/factsheets/fs104/en/index.html> (2012).
- Russell, D. G., Barry, C. E. III & Flynn, J. L. Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**, 852–856 (2010).
- Bishai, W., Sullivan, Z., Bloom, B. R. & Andersen, P. Bettering BCG: a tough task for a TB vaccine? *Nature Med.* **19**, 410–411 (2013).
- Lawn, S. D. & Zumla, A. I. Tuberculosis. *Lancet* **378**, 57–72 (2011).
- Gandhi, N. R. et al. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* **375**, 1830–1843 (2010).
- Nathan, C. Fresh approaches to anti-infective therapies. *Sci. Transl. Med.* **4**, 140sr142 (2012).
- Kaufmann, S. H. & Dorhoi, A. Inflammation in tuberculosis: interactions, imbalances and interventions. *Curr. Opin. Immunol.* **25**, 441–449 (2013).
- Mayer-Barber, K. D. et al. Innate and adaptive interferons suppress IL-1 α and IL-1 β production by distinct pulmonary myeloid subsets during *Mycobacterium tuberculosis* infection. *Immunity* **35**, 1023–1034 (2011).
- Mayer-Barber, K. D. et al. Caspase-1 independent IL-1 β production is critical for host resistance to *Mycobacterium tuberculosis* and does not require TLR signaling in vivo. *J. Immunol.* **184**, 3326–3330 (2010).
- Serhan, C. N., Chiang, N. & Van Dyke, T. E. Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nature Rev. Immunol.* **8**, 349–361 (2008).
- Bafica, A. et al. Host control of *Mycobacterium tuberculosis* is regulated by 5-lipoxygenase-dependent lipoxin production. *J. Clin. Invest.* **115**, 1601–1606 (2005).
- Chen, M. et al. Lipid mediators in innate immunity against tuberculosis: opposing roles of PGE2 and LXA4 in the induction of macrophage death. *J. Exp. Med.* **205**, 2791–2801 (2008).
- Divangahi, M. et al. *Mycobacterium tuberculosis* evades macrophage defenses by inhibiting plasma membrane repair. *Nature Immunol.* **10**, 899–906 (2009).
- Kirtikara, K. et al. Compensatory prostaglandin E2 biosynthesis in cyclooxygenase 1 or 2 null cells. *J. Exp. Med.* **187**, 517–523 (1998).
- Antonelli, L. R. et al. Intranasal Poly-IC treatment exacerbates tuberculosis in mice through the pulmonary recruitment of a pathogen-permissive monocyte/macrophage population. *J. Clin. Invest.* **120**, 1674–1682 (2010).
- Rayamajhi, M., Humann, J., Kearney, S., Hill, K. K. & Lenz, L. L. Antagonistic crosstalk between type I and II interferons and increased host susceptibility to bacterial infections. *Virulence* **1**, 418–422 (2010).
- Manca, C. et al. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN- α/β . *Proc. Natl Acad. Sci. USA* **98**, 5752–5757 (2001).
- Stanley, S. A., Johndrow, J. E., Manzanillo, P. & Cox, J. S. The type I IFN response to infection with *Mycobacterium tuberculosis* requires ESX-1-mediated secretion and contributes to pathogenesis. *J. Immunol.* **178**, 3143–3152 (2007).
- O'Garra, A. et al. The immune response in tuberculosis. *Annu. Rev. Immunol.* **31**, 475–527 (2013).
- Berry, M. P. et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973–977 (2010).
- Maertzdorf, J. et al. Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun.* **12**, 15–22 (2011).
- Ottenshoff, T. H. et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS ONE* **7**, e45839 (2012).
- Barry, C. E. III et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Rev. Microbiol.* **7**, 845–855 (2009).
- Law, K. et al. Increased release of interleukin-1 beta, interleukin-6, and tumor necrosis factor-alpha by bronchoalveolar cells lavaged from involved sites in pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **153**, 799–804 (1996).
- Tsao, T. C. et al. Increased TNF-alpha, IL-1 beta and IL-6 levels in the bronchoalveolar lavage fluid with the upregulation of their mRNA in macrophages lavaged from patients with active pulmonary tuberculosis. *Tuber. Lung Dis.* **79**, 279–285 (1999).
- Kuo, H. P. et al. Nitric oxide modulates interleukin-1 beta and tumor necrosis factor-alpha synthesis by alveolar macrophages in pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **161**, 192–199 (2000).
- Wang, C. H. & Kuo, H. P. Nitric oxide modulates interleukin-1 beta and tumor necrosis factor-alpha synthesis, and disease regression by alveolar macrophages in pulmonary tuberculosis. *Respirology* **6**, 79–84 (2001).
- Rousu, J., Agranoff, D. D., Sodeinde, O., Shawe-Taylor, J. & Fernandez-Reyes, D. Biomarker discovery by sparse canonical correlation analysis of complex clinical phenotypes of tuberculosis and malaria. *PLOS Comput. Biol.* **9**, e1003018 (2013).
- Tobin, D. M. et al. The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717–730 (2010).
- Tobin, D. M. et al. Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* **148**, 434–446 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the NIAID Intramural Research program and a Concept Acceleration Program-Award (K.D.M.-B., B.B.A. and A.S.) from DMID, NIAID. We are grateful to K. Elkins, S. Morris, M. Belcher as well as the NIAID ABSL3 support staff for facilitating our animal studies. We thank R. Chen, L. Goldfeder and Q. Gao for sharing their clinical trial expertise and research facilities, respectively. We also thank K. Kauffman, R. Thompson, S. Hieny, P. Dayal, D. Surman, L. Meng, Z. Li, L. Lifa, Q. Shen and Z. Huang for technical assistance, H. Boshoff for help with direct anti-mycobacterial activity assays and M. S. Jawahar, V. V. Banurekha and R. Sridhar for recruitment and clinical evaluation of patients in Chennai, India. We are grateful to F. Andrade Neto, H. Remold, K. Arora, J. Aliberti, M. Moayeri, P. Murphy, A. O'Garra, R. Germain and C. Serhan for discussion or critical reading of the manuscript. Finally, we thank the patients, volunteer participants, and clinical staff of the Tuberculosis department of Henan Chest Hospital in Zhengzhou, China and the Department of Clinical Research (NIRT) and Department of Thoracic Medicine (Government Stanley Medical Hospital) in Chennai, India for their participation in our clinical studies.

Author Contributions K.D.M.-B. conceived the study, designed and performed experiments, analysed data and wrote the paper; B.B.A. performed experiments, analysed data and prepared the Indian cohort description; E.P.A. and D.L.B. performed experiments; S.D.O., J.G., S.C.D., N.P.K., Y.C., L.E.V., provided technical or analytical assistance; S.B. recruited, sampled and collected data about patients and provided access to samples from Indian cohort, M.C. provided healthy donor material, A.M.S. provided Hiltanol (pCLC); R.S., W.W., X.Y., G.Z., L.E.V. and C.E.B. conducted the Natural History Study in Zhengzhou, provided access to Chinese patient samples and the preparation of the Chinese cohort description, A.S. provided conceptual advice and wrote the paper and all authors approved the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.D.M.-B. (mayerk@niaid.nih.gov).

Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors

Johannes Popow^{1,2}, Jennifer Jurkin¹, Alexander Schleiffer³ & Javier Martinez¹

RNA ligases have essential roles in many cellular processes in eukaryotes, archaea and bacteria, including in RNA repair^{1,2} and stress-induced splicing of messenger RNA³. In archaea and eukaryotes, RNA ligases also have a role in transfer RNA splicing to generate functional tRNAs required for protein synthesis^{4–7}. We recently identified the human tRNA splicing ligase, a multimeric protein complex with RTCB (also known as HSPC117, C22orf28, FAAP and D10Wsu52e) as the essential subunit⁸. The functions of the additional complex components ASW (also known as C2orf49), CGI-99 (also known as C14orf166), FAM98B and the DEAD-box helicase DDX1 in the context of RNA ligation have remained unclear. Taking advantage of clusters of eukaryotic orthologous groups, here we find that archease (ARCH; also known as ZBTB80S), a protein of unknown function, is required for full activity of the human tRNA ligase complex and, in cooperation with DDX1, facilitates the formation of an RTCB–guanylate intermediate central to mammalian RNA ligation. Our findings define a role for DDX1 in the context of the human tRNA ligase complex and suggest that the widespread co-occurrence of archease and RtcB proteins implies evolutionary conservation of their functional interplay.

A kinetic assay using Flag-tagged RTCB affinity purified from stably transfected HEK293 cells (Extended Data Fig. 1a, b) and tRNA exon halves (Fig. 1a) or isolated linear intron as a minimal substrate (Fig. 1b, c) (see Methods) revealed that product formation virtually ceased to occur before consumption of the substrate (Fig. 1a–c). The maximum amount of ligation product formed was proportional to the initial amount of

enzyme (Fig. 1c and Extended Data Fig. 1c–e), suggesting that affinity-purified Flag–RTCB requires an additional, stimulatory component for unlimited enzymatic turnover. We reasoned that the distinctive phyletic distribution of RtcB proteins in eukaryotes (that is, their absence in plants and fungi⁹, which rely on different ligase proteins and mechanisms^{10–12}; Extended Data Fig. 2a) could be used to discover proteins functionally associated with human RTCB. We therefore focused our attention on seven clusters of eukaryotic orthologous groups (KOGs) conserved in the same model organisms as RtcB (cluster number KOG3833)^{13,14} (Extended Data Fig. 2b). One of these, KOG4528, encompasses archease, a protein of unknown function predicted to act in combination with nucleic acid processing factors¹⁵. In *Pyrococcus abyssi*, archease improves the substrate specificity of a tRNA methyltransferase encoded downstream in a bicistronic operon¹⁶; however, this function does not appear to be conserved in mammals¹⁷.

We therefore added recombinant archease (Extended Data Fig. 3a) to ligase assays containing affinity-purified human Flag–RTCB, which made ligase reactions proceed to near-complete consumption of substrate (Fig. 1d). Addition of recombinant archease to stalled ligase reactions allowed the formation of product to resume (Fig. 1e and Extended Data Fig. 3b, c). Two mutant versions of archease, D39A (Asp 39 → Ala 39) and K144A (Lys 144 → Ala 144) (Extended Data Fig. 4), had no effect on the ligase activity of human Flag–RTCB (Fig. 1f).

Affinity purification of Flag–archease or Flag–RTCB from HEK293 cells did not reveal a detectable association of archease with the human tRNA ligase complex (Fig. 2a) or any RNA ligase activity (Fig. 2b). However,

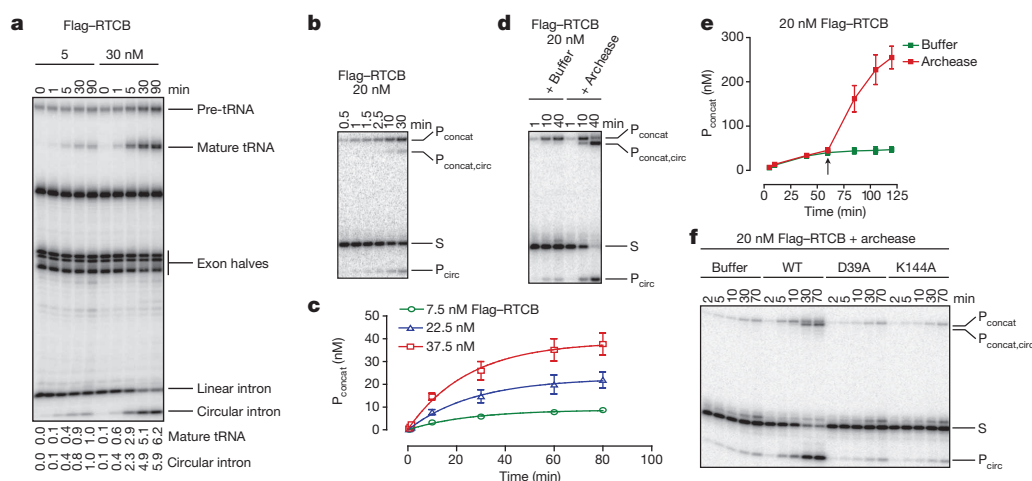


Figure 1 | Archease facilitates multiple enzymatic turnover of RTCB.

a, Flag–RTCB (5 or 30 nM) was incubated with cleaved pre-tRNA transcripts. Numbers below lanes indicate amounts of mature tRNA and intron (arbitrary numbers relative to 5 nM Flag–RTCB at 90 min, set to 1.0). **b**, Flag–RTCB converts linear RNA fragments (S, 100 nM) into concatemized (P_{concat}), circularized concatemer ($P_{\text{concat,circ}}$) and circularized (P_{circ}) products. **c**, The endpoint of ligation reactions (1 μM linear RNA substrate) depends on

Flag–RTCB concentration. Values represent mean and standard deviation (s.d.) of technical replicates ($N = 3$). **d**, RNA ligase assays (100 nM substrate) in the presence or absence of recombinant archease (10 μM). **e**, Addition of archease (10 μM , arrow) restarts stalled ligase reactions (500 nM substrate) (mean \pm s.d., $N = 3$). **f**, Ligase reactions in the presence of buffer, recombinant wild-type (WT), D39A- or K144A-mutant archease (10 μM) with linear substrate (1 μM).

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), A-1030 Vienna, Austria. ²European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. ³IMP/IMBA Bioinformatics Core Facility, Research Institute of Molecular Pathology (IMP), A-1030 Vienna, Austria.

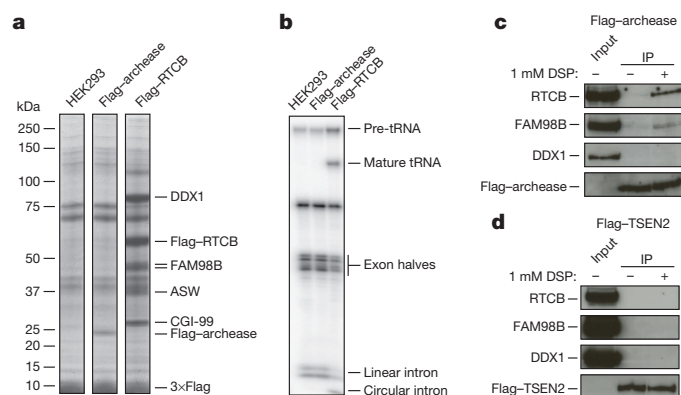


Figure 2 | Archease co-purifies with human RTCB upon *in vivo* crosslink. **a**, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Coomassie blue) of Flag-archaease and Flag-RTCB affinity purified from HEK293 cells (150 mM NaCl). **b**, RNA ligase assay (cleaved pre-tRNA transcripts) of Flag-archaease and Flag-RTCB preparations analysed in **a**. **c**, Western blot of Flag-archaease affinity purified from HEK293 cells treated with 1 mM DSP or dimethylsulphoxide (DMSO). IP, immunoprecipitates. **d**, Western blot of Flag-TSEN2 affinity purified from HEK293 cells treated with 1 mM DSP or DMSO.

addition of the cell-permeable crosslinking reagent dithiobis(succinimidylpropionate) (DSP) to HEK293 cells before affinity purification enabled us to detect a specific association of Flag-archaease with endogenous RTCB and FAM98B (Fig. 2c and Extended Data Fig. 5a, b). Failure to detect an association with Flag-TSEN2 (ref. 18), which is not stimulated

by archease (Extended Data Fig. 5c, d), further attests to the specificity of the crosslink (Fig. 2d). Thus, archease is a component of the human tRNA ligase complex. Characterization of the interaction between archease and the tRNA ligase complex by surface plasmon resonance (SPR) yielded a dissociation constant (K_d) of 1.8 ± 0.9 nM (Extended Data Fig. 5e–g). The inactive D39A mutant had similar interaction parameters to the wild-type protein, indicating that this residue directly mediates the chemical reaction steps or conformational rearrangements required for RNA ligation rather than facilitating the interaction of archease with RTCB.

To test whether archease stimulates the activity of human RTCB complexes at the stage of product release, we established a gel shift assay to detect RTCB–RNA adducts forming at reaction conditions. These adducts (termed C1, C2 and C2a) appeared only transiently in the presence of wild-type archease (Fig. 3a, lanes 1–4) but persisted upon mutagenesis of conserved residues (Fig. 3a, lanes 5–8, and Extended Data Fig. 6a). All adducts—irrespective of whether they had formed in the presence or absence of functional archease—exclusively contained the substrate RNA species S and P_{concat} (Fig. 3b, lanes 1–4, and Extended Data Fig. 6b) and Flag-RTCB (Extended Data Fig. 6c). Taken together, these data do not support the release of RNA products being a rate-limiting step accelerated by archease.

The ligase activity of RtcB from *Escherichia coli* depends on the formation of a covalent RtcB–guanylate intermediate and the subsequent transfer of GMP to the 3' end of RNA molecules with 2',3'- or 3'-phosphate termini^{19,20}. In contrast to prokaryotic RtcB, addition of [α -³²P] GTP to RNA ligase assays did not lead to the formation of a radiolabelled human RTCB–guanylate adduct (Fig. 3c, top, lane 1). Formation of a radiolabelled band corresponding to guanylated human Flag-RTCB

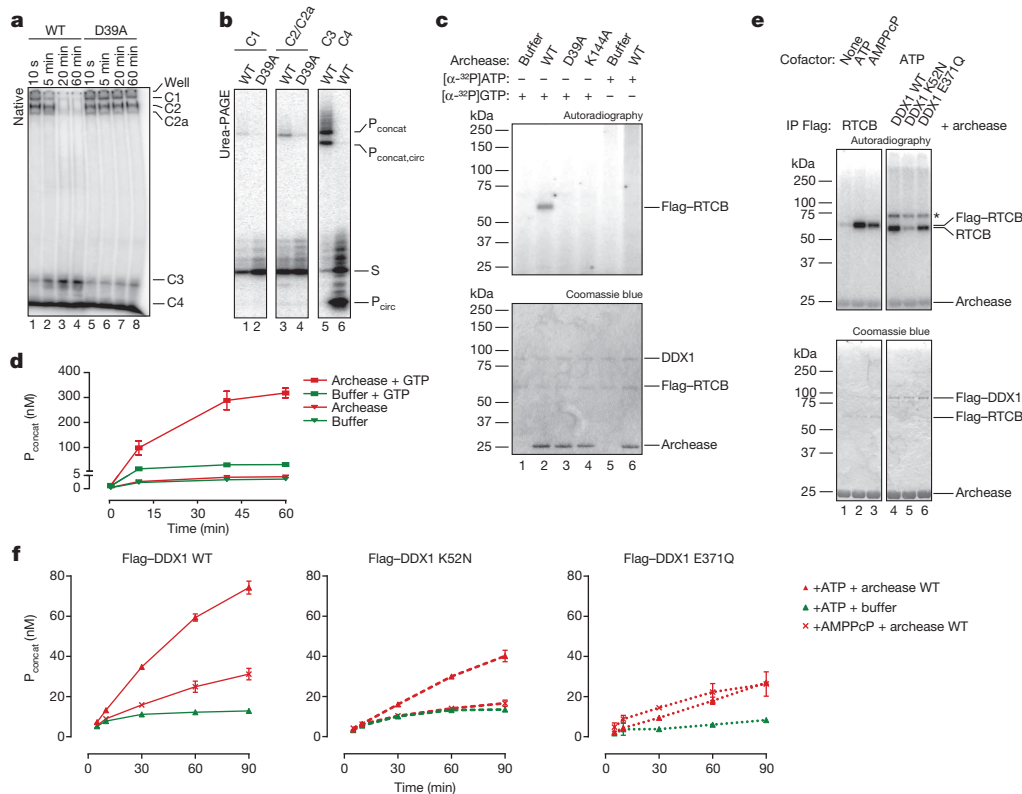


Figure 3 | Archease cooperates with DDX1 to guanylate RTCB *in vitro*. **a**, Time-course native PAGE of Flag-RTCB–RNA adducts in the presence of wild-type (WT) or D39A mutant archease. **b**, Denaturing PAGE of the RNA content of Flag-RTCB–RNA adducts at 5 min in the presence of wild-type or mutant archease. **c**, SDS-PAGE of Flag-RTCB incubated with buffer, wild-type or mutant (D39A, K144A) archease and [α -³²P]GTP or [α -³²P]ATP. **d**, Ligation reactions (20 nM Flag-RTCB, 500 nM substrate) in the presence of buffer or archease (10 μ M) omitting (triangles) or including GTP (0.5 mM,

squares) (mean \pm s.d., $N = 3$). **e**, SDS-PAGE of Flag-RTCB, wild-type or mutant Flag-DDX1 (K52N, E371Q) incubated with 0.1 mM EDTA, ATP or AMPPcP, [α -³²P]GTP and archease. The asterisk indicates an unrelated band. **f**, Ligase assays (1 μ M substrate) of 20 nM wild-type or mutant (K52N, E371Q) Flag-DDX1 in the presence (red triangles and crosses) or absence (green triangles) of archease (10 μ M) and 0.5 mM ATP (red and green triangles) or AMPPcP (red crosses) (mean \pm s.d., $N = 3$).

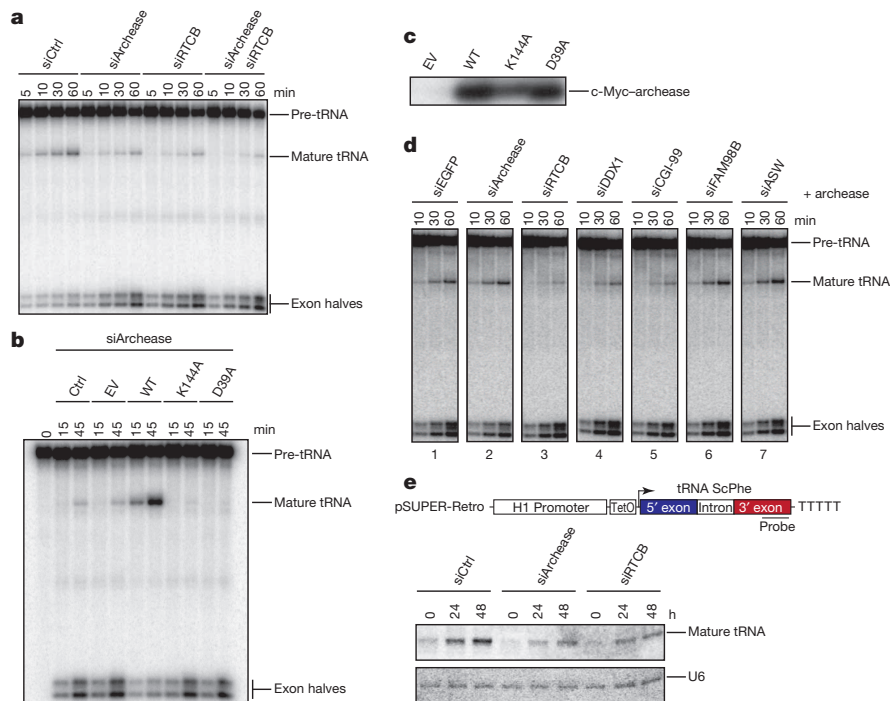


Figure 4 | tRNA maturation relies on RTCB, archease and DDX1 *in vitro* and in living cells. **a**, *In vitro* tRNA maturation assay of HeLa extracts that were RNAi depleted of archease, RTCB or both. Ctrl, control. **b**, *In vitro* tRNA maturation assay of HeLa extracts that were RNAi depleted of archease and simultaneously transfected with vector (EV), c-Myc-tagged wild-type or mutant (D39A, K144A) archease. **c**, Western blot of extracts assayed in **b**.

required the addition of wild-type archease, whereas inactive archease point mutants did not support guanylation (Fig. 3c, top, lanes 2–4). In agreement with this finding, archease stimulated the activity of the human tRNA ligase complex with high efficiency only in the presence of GTP (Fig. 3d and Extended Data Fig. 7a). The co-purification of GTP or isolation of a guanylated form of RTCB during affinity purification^{8,20} may account for the detection of activity corresponding to incomplete turnover cycles during the initial and apparently archease-independent phase of the reaction (Extended Data Fig. 7a). Although ATP could not serve as a nucleotidyl donor for human RTCB (Fig. 3c, lanes 5 and 6), we observed a marked stimulation of guanylation by ATP in the presence of archease (Fig. 3e, compare lanes 1 and 2) and therefore tested whether guanylation of RTCB depends on the ATP-binding DEAD-box helicase DDX1 (Extended Data Fig. 7b–f). Both substitution of ATP by its non-hydrolysable ATP analogue AMPPcP and mutagenesis of two conserved motifs of DDX1 implicated in ATP binding and hydrolysis²¹ (Extended Data Fig. 7b, d, e) affected archease-dependent guanylation (Fig. 3e, compare lane 2 with lane 3, and lane 4 with lanes 5 and 6), indicating that ATP hydrolysis by DDX1 is required for maximal stimulation of the human RNA ligase by archease. RTCB guanylation was most compromised for DDX1 K52N, the mutant with the lowest affinity for ATP and RNA (Fig. 3e, lane 5, and Extended Data Figs 7e, f and 8a). In agreement with these results, RTCB complexes with mutagenized DDX1 exhibited diminished RNA ligase activity (Fig. 3f and Extended Data Figs 8b–e and 9), thus providing a molecular basis for the dual cofactor requirement of human RNA ligase.

Depletion of archease by RNA interference (RNAi) impaired maturation of intron containing pre-tRNAs *in vitro*, to a comparable extent as observed upon the depletion of RTCB (Fig. 4a and Extended Data Fig. 10a, b). Simultaneous depletion of archease and RTCB suppressed pre-tRNA maturation even more (Fig. 4a). Splicing activity was restored by expressing exogenous wild-type but not mutant archease (Fig. 4b, c). In agreement with the dependence of the activity of the human tRNA

ligase complex on DDX1, tRNA maturation under multiple turnover conditions—that is, in the presence of recombinant archease—was recognizably impaired upon its depletion. A similar effect was observed upon the depletion of CGI-99, presumably due to the concomitant depletion of DDX1 (Fig. 4d, compare panels 1 and 2 with panels 4 and 5, and Extended Data Fig. 10c)⁸. Induction of transcription of tagged pre-tRNAs⁸ revealed that the formation of mature tRNA was impaired to comparable levels in cells depleted of archease or RTCB (Fig. 4e and Extended Data Fig. 10d).

We have revealed that archease facilitates the DDX1-dependent formation of an RTCB–guanylate intermediate central to mammalian RNA ligation and provided evidence for the role of ATP hydrolysis by the human tRNA ligase^{8,22} (Extended Data Fig. 7a). In human cell lines, a major fraction of DDX1 associates with the tRNA ligase complex, which is supported by affinity purification (Extended Data Fig. 7c) and short interfering (si)RNA-mediated co-depletion experiments (Extended Data Fig. 10c). These findings suggest that RTCB—and possibly archease—may also be involved in the established functions of DDX1 in the context of retinoblastoma²³, cytoplasmic ribonucleoprotein particles²⁴, viral replication^{25,26} and the induction of antiviral responses in dendritic cells²⁷. Our implication of archease in mammalian tRNA splicing has already stimulated experiments demonstrating the functional cooperation of archease and RtcB proteins in at least one archaeon²⁸, and this may be widespread as many prokaryotic genomes harbour homologues of both proteins in operons^{29,30}.

METHODS SUMMARY

Flag-tagged RTCB, archease, TSEN2 and DDX1 were isolated by affinity chromatography from stably transfected HEK293 cells. RNA ligase and pre-tRNA processing assays were carried out as described previously⁸. KOGs were retrieved from the publically available database [ftp://ftp.ncbi.nih.gov/pub/COG/KOG/kog](http://ftp.ncbi.nih.gov/pub/COG/KOG/kog)^{13,14}. SPR measurements were carried out and analysed by F. Krick at NBS-C BioScience & Consulting GmbH. A detailed description of all the methods and reagents used can be found in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 December 2013; accepted 27 March 2014.

Published online 25 May 2014.

- Chan, C. M., Zhou, C. & Huang, R. H. Reconstituting bacterial RNA repair and modification *in vitro*. *Science* **326**, 247 (2009).
- Nandakumar, J., Schwer, B., Schaffrath, R. & Shuman, S. RNA repair: an antidote to cytotoxic eukaryal RNA damage. *Mol. Cell* **31**, 278–286 (2008).
- Sidrauski, C., Cox, J. S. & Walter, P. tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. *Cell* **87**, 405–413 (1996).
- Filipowicz, W. & Shatkin, A. J. Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. *Cell* **32**, 547–557 (1983).
- Konarska, M., Filipowicz, W. & Gross, H. J. RNA ligation via 2'-phosphomonoester, 3'5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase. *Proc. Natl Acad. Sci. USA* **79**, 1474–1478 (1982).
- Laski, F. A., Fire, A. Z., RajBhandary, U. L. & Sharp, P. A. Characterization of tRNA precursor splicing in mammalian extracts. *J. Biol. Chem.* **258**, 11974–11980 (1983).
- Zofalova, L., Guo, Y. & Gupta, R. Junction phosphate is derived from the precursor in the tRNA spliced by the archaeon *Haloferax volcanii* cell extract. *RNA* **6**, 1019–1030 (2000).
- Popow, J. *et al.* HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science* **331**, 760–764 (2011).
- Popow, J., Schleiffer, A. & Martinez, J. Diversity and roles of (t)RNA ligases. *Cell. Mol. Life Sci.* **69**, 2657–2670 (2012).
- Englert, M. & Beier, H. Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Res.* **33**, 388–399 (2005).
- Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature* **293**, 112–116 (1981).
- Phizicky, E. M., Consaul, S. A., Nehrke, K. W. & Abelson, J. Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. *J. Biol. Chem.* **267**, 4577–4582 (1992).
- Koonin, E. V. *et al.* A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
- Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41 (2003).
- Canaves, J. M. Predicted role for the archease protein family based on structural and sequence analysis of TM1083 and MTH1598, two proteins structurally characterized through structural genomics efforts. *Proteins* **56**, 19–27 (2004).
- Auxilien, S., El Khadali, F., Rasmussen, A., Douthwaite, S. & Grosjean, H. Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m⁵C methyltransferase. *J. Biol. Chem.* **282**, 18711–18721 (2007).
- Brzezicha, B. *et al.* Identification of human tRNA:m⁵C methyltransferase catalysing intron-dependent m⁵C formation in the first position of the anticodon of the pre-tRNA Leu (CAA). *Nucleic Acids Res.* **34**, 6034–6043 (2006).
- Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. *Cell* **117**, 311–321 (2004).
- Chakravarty, A. K., Subbotin, R., Chait, B. T. & Shuman, S. RNA ligase RtcB splices 3'-phosphate and 5'-OH ends via covalent RtcB-(histidiny)-GMP and polynucleotide-(3')pp(5')G intermediates. *Proc. Natl Acad. Sci. USA* **109**, 6072–6077 (2012).
- Tanaka, N., Chakravarty, A. K., Maughan, B. & Shuman, S. A novel mechanism of RNA repair by RtcB via sequential 2',3'-cyclic phosphodiesterase and 3'-phosphate/5'-hydroxyl ligation reactions. *J. Biol. Chem.* **286**, 43134–43143 (2011).
- Pause, A. & Sonenberg, N. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11**, 2643–2654 (1992).
- Perkins, K. K., Furneaux, H. & Hurwitz, J. Isolation and characterization of an RNA ligase from HeLa cells. *Proc. Natl Acad. Sci. USA* **82**, 684–688 (1985).
- Godbout, R., Packer, M. & Bie, W. Overexpression of a DEAD box protein (DDX1) in neuroblastoma and retinoblastoma cell lines. *J. Biol. Chem.* **273**, 21161–21168 (1998).
- Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513–525 (2004).
- Cao, D., Haussecker, D., Huang, Y. & Kay, M. A. Combined proteomic–RNAi screen for host factors involved in human hepatitis delta virus replication. *RNA* **15**, 1971–1979 (2009).
- Edgcomb, S. P. *et al.* DDX1 is an RNA-dependent ATPase involved in HIV-1 Rev function and virus replication. *J. Mol. Biol.* **415**, 61–74 (2012).
- Zhang, Z. *et al.* DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity* **34**, 866–878 (2011).
- Desai, K. K., Cheng, C. L., Bingman, C. A., Phillips, G. N. Jr & Raines, R. T. A tRNA splicing operon: archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation. *Nucleic Acids Res.* (2014).
- Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
- Ciria, R., Abreu-Goodger, C., Morett, E. & Merino, E. GeConT: gene context analysis. *Bioinformatics* **20**, 2307–2308 (2004).

Acknowledgements We would like to thank S. Ameres, B. Beckmann, R. Hartmann, M. Hentze, M. Jantsch, B. Klaus, B. Mair, A. F. Nielsen, J.-M. Peters, T. Rapoport, R. Schroeder and S. Weitzer for help and advice, S. Bandini for the preparation of Flag-TSEN2 cell lines and critical discussions and J. Dammann and T. Lendl for technical assistance. This work has been funded by the Fonds zur Förderung der wissenschaftlichen Forschung (P24687), the GEN-AU 3 research programme (820982 Non-coding RNAs) (J.P.) and the Institute of Molecular Biotechnology of the Austrian Academy of Sciences (J.M.).

Author Contributions J.P. designed and carried out experiments and wrote the manuscript, J.J. designed and carried out experiments and contributed to writing the manuscript, A.S. performed bioinformatic analysis of archease and contributed to writing the manuscript, J.M. designed the experiments and contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (javier.martinez@imba.oeaw.ac.at).

Structural basis for lipopolysaccharide insertion in the bacterial outer membrane

Shuai Qiao^{1,2}, Qingshan Luo^{1,2}, Yan Zhao^{1,3}, Xuejun Cai Zhang¹ & Yihua Huang¹

One of the fundamental properties of biological membranes is the asymmetric distribution of membrane lipids. In Gram-negative bacteria, the outer leaflet of the outer membrane is composed predominantly of lipopolysaccharides (LPS)¹. The export of LPS requires seven essential lipopolysaccharide transport (Lpt) proteins to move LPS from the inner membrane, through the periplasm to the surface². Of the seven Lpt proteins, the LptD–LptE complex is responsible for inserting LPS into the external leaflet of the outer membrane^{3,4}. Here we report the crystal structure of the ~110-kilodalton membrane protein complex LptD–LptE from *Shigella flexneri* at 2.4 Å resolution. The structure reveals an unprecedented two-protein plug-and-barrel architecture with LptE embedded into a 26-stranded β-barrel formed by LptD. Importantly, the secondary structures of the first two β-strands are distorted by two proline residues, weakening their interactions with neighbouring β-strands and creating a potential portal on the barrel wall that could allow lateral diffusion of LPS into the outer membrane. The crystal structure of the LptD–LptE complex opens the door to new antibiotic strategies targeting the bacterial outer membrane.

The outer membrane of Gram-negative bacteria is comprised of an asymmetric lipid bilayer with phospholipids in the inner leaflet and glycolipids, predominately lipopolysaccharide (LPS), in the outer leaflet¹. The LPS layer in the outer membrane endows Gram-negative bacteria with a strong permeability barrier against toxic compounds such as antibiotics, allowing survival in harsh environments⁵. LPS is also a potent activator of the innate immune response and acts as a conserved pathogen-associated molecular pattern (PAMP) recognized by innate immune receptors⁶. Given its functional importance, it is not surprising that LPS is essential to most Gram-negative bacteria, and intervening in its biogenesis pathway offers great opportunities for developing novel antibiotics against pathogenic bacteria.

In *Escherichia coli*, the transport of LPS molecules from the outer leaflet of the inner membrane to the cell surface is executed by the LPS transport (Lpt) machinery⁷, which consists of seven essential proteins (LptA–LptG)². Once extracted from the inner membrane by an ATP-binding-cassette (ABC) transporter complex comprised of LptB, LptF and LptG⁸, the detached LPS molecule is delivered to LptC⁹ and further to LptA in an ATP-dependent manner¹⁰. The final stage of LPS transport to the cell surface is accomplished by an outer-membrane-localized two-protein complex, LptD–LptE^{3,4}. However, the precise mechanism of LPS insertion and assembly in the outer membrane remains obscure. Here, we report the crystal structure of the LptD–LptE complex from *Shigella flexneri* at 2.4 Å resolution. The structure provides mechanistic insights into how LPS is exported to and inserted in the outer leaflet of the outer membrane.

One remarkable feature of the LptD–LptE complex is the observation of two pairs of non-consecutive disulphide bonds in LptD: residues Cys 31 and Cys 173 form disulphide bonds with residues Cys 724 and Cys 725, respectively, confirming previous functional analyses^{11–13} (Fig. 1a, c). To validate the presence of disulphide bonds in the purified recombinant protein of LptD, we performed SDS–PAGE analysis for purified

LptD–LptE complex under both reducing and non-reducing conditions. As shown in Fig. 1b, LptD exhibited slower mobility on SDS–PAGE with an apparent molecular weight of 130 kDa under non-reducing conditions, but returned to its calculated molecular weight of 87 kDa under reducing conditions, indicating the presence of disulphide bonds within LptD. The LptD–LptE complex was unable to dissociate on SDS–PAGE

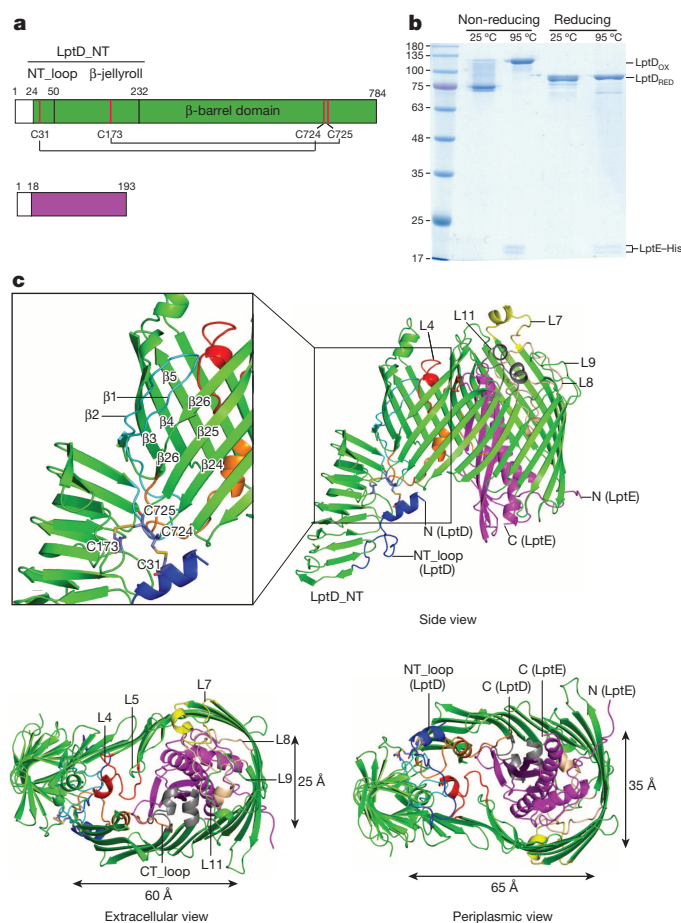


Figure 1 | General architecture of the LptD–LptE complex. **a**, Schematic structures of LptD and LptE. **b**, SDS–PAGE showing the presence of disulphide bonds within LptD and stability of the LptD–LptE complex. LptD_{OX} and LptD_{RED} represent the oxidized and reduced LptD, respectively. Data shown are representative of at least three independent experiments. **c**, Cartoon representation of the LptD–LptE complex. LptD and LptE are coloured in green and magenta, respectively. Strands β1 and β2 are highlighted in cyan. Extracellular loops L4, L7, L8 and L11 are coloured in red, yellow, wheat and grey, respectively. The N-terminal loop (NT_loop) and the C-terminal loop (CT_loop) of LptD are highlighted in blue and orange, respectively. Dimensions of the LptD–LptE complex are marked.

¹National Laboratory of Biomacromolecules, National Center of Protein Science–Beijing, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ²University of Chinese Academy of Sciences, Beijing 100101, China. ³School of Life Sciences, University of Science and Technology of China, Hefei 230027, Anhui, China.

unless the samples were heated for 10 min at 95 °C, suggesting LptD and LptE form a tight complex (Fig. 1b). Additionally, disruption of any of the two disulphide bonds caused partial degradation of LptD and promiscuous disulphide bond formation within LptD when co-expressed with LptE (Extended Data Fig. 3).

The amino terminus of LptD (LptD_NT, residues 24–232) can be subdivided into a 26-residue fragment (NT_loop, residues 24–50) and a β -jellyroll domain (Figs 1c and 2a), confirming a previous prediction². Each of the two β -sheets of the β -jellyroll consists of eleven antiparallel β -strands^{14,15} (Fig. 2a). The interior surface of the β -jellyroll is highly hydrophobic with two bound detergent molecules (Fig. 2a and Extended Data Fig. 7a), indicating its role in binding the hydrophobic moiety of LPS in a manner that resembles LptA and LptC^{16,17}. Structure superposition of LptD_NT with LptA and LptC showed high structural similarities between them, with an r.m.s.d. (root-mean-square deviation) of 1.7 Å (117 C α atoms) to LptA and 2.6 Å (106 C α atoms) to LptC, respectively (Fig. 2b, c). Previous studies showed that residues Ile 36, Phe 95 and Leu 116 of LptA bind LPS *in vivo*¹⁰. The high structural similarity between LptD_NT and LptA implies that corresponding residues Val 51, Tyr 112 and Leu 128 of LptD_NT may also bind LPS (Fig. 2b and Extended Data Fig. 4). The NT_loop of LptD_NT starts with a short α -helix near the β -barrel and runs in the reverse direction from that of the β -jellyroll (Figs 1c and 2a).

The carboxy terminus of LptD (residues 233–784) forms a kidney-shaped β -barrel composed of twenty-six antiparallel β -strands (termed β 1– β 26; Fig. 1c). The β -barrel has periplasmic and extracellular dimensions of 65 Å \times 35 Å and 60 Å \times 25 Å, respectively (Fig. 1c, C α distance). Therefore, LptD is the largest single-protein β -pore observed to date. The plugging of LptE inside of the barrel causes a diminished lumen size to 45 Å \times 35 Å on the periplasmic side. Given that the dimensions of Ra LPS molecule is about 32 Å \times 28 Å \times 12 Å (ref. 18; Extended Data Fig. 1b), the size of the vacant lumen inside of the LptD–LptE complex on the periplasmic side is sufficiently large for accommodating Ra LPS.

Strikingly, among the 26 strands that constitute the barrel, the secondary structures of the first two strands are distorted, presumably because of the presence of two proline residues at the positions of 231 and 246 (Fig. 3a). In particular, Pro 231 terminates the strand β 1 at the N-terminal end. As a consequence, β 1 forms only three main-chain hydrogen bonds to its neighbouring strand β 26 (Fig. 3b). Likewise, Pro 246 breaks strand β 2 in the middle, leaving only four and five hydrogen bonds to β 1 and β 3, respectively (Fig. 3b). The interaction between β 1 and β 2 is also mediated by an inter-strand salt bridge formed between Glu 242 and Lys 234 inside the β -barrel (Extended Data Fig. 8). Additionally, Pro 261, which is located in strand β 3, may also contribute to weakened β -sheet formation with the upper portion of β 2 (Fig. 3a). An alignment of LptD orthologues from 25 species revealed the conservation of proline residues, Pro 231, Pro 246 and Pro 261 (Extended Data Fig. 2). We predict that the hairpin formed by strands β 1 and β 2 may function as a door, creating a gate with a width as large as 16 Å between strands β 3 and β 26 upon dislocation (Fig. 3a). Consistent with this hypothesis, the average B-factors for both β 1 and upper portion of β 2 (36.8 Å² for C α atoms of residues 228–249) are much higher than those of other β -strands (21.0 Å² for C α atoms of residues 255–757) that constitute the rest of the β -barrel, indicating that this portion is more dynamic in the membrane. Intriguingly, side chains of four aromatic residues (Tyr 235, Phe 241, Tyr 244 and Phe 754) adopt alternating conformations and are all located at the potential LPS exit portal (Fig. 3a and Extended Data Fig. 7b). It is likely that these aromatic residues may play certain roles in LPS export into the outer membrane.

Unlike the highly hydrophobic nature of the interiors of LptA, LptC, LptD_NT and MD-2, which all bind the hydrophobic moiety of LPS^{14,15,18}, the interior of the LptD–LptE complex is fairly hydrophilic (Extended Data Fig. 5). The charged residues do not form clusters but are uniformly distributed inside the barrel, suggesting that the entry of the hydrophobic moiety of LPS to the barrel is energetically unfavourable, but this inner surface property may favour the hydrophilic moiety of

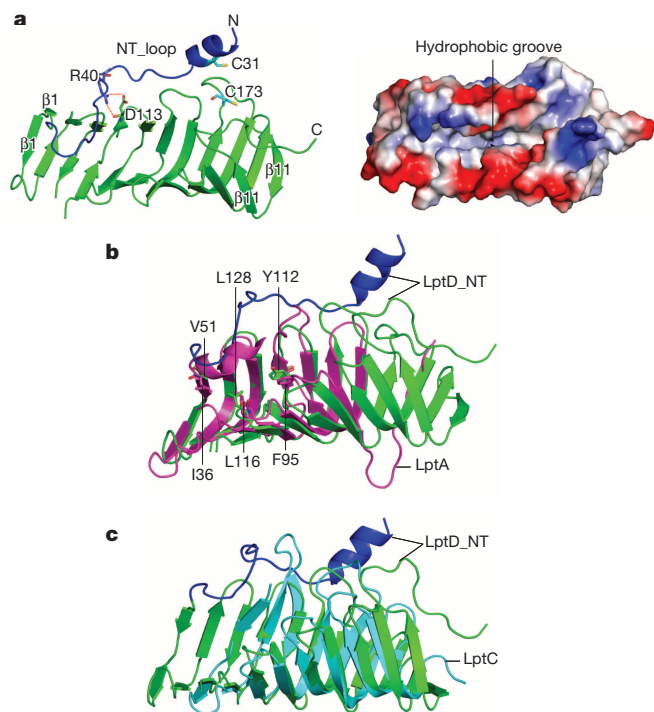


Figure 2 | The N terminus of LptD (LptD_NT) forms a β -jellyroll that functions in LPS transport. **a**, Cartoon and electrostatic surface representation of LptD_NT. LptD_NT contains an NT_loop (blue) and a β -jellyroll (green). Surface representation of the electrostatic potential (right) shows the hydrophobic interior of the β -jellyroll. For clarity, the NT_loop is omitted. **b**, Superposition of LptD_NT (green) with LptA (magenta) showing the structural similarities. Residues V51, Y112 and L128 of LptD_NT may bind the fatty acyl chains of LPS. **c**, Superposition of LptD_NT (green) with LptC (cyan).

LPS entry into the barrel of the LptD–LptE complex. Consistent with this hypothesis, there is a large hole at the interface between LptD_NT and the LptD barrel, and a portion of LptD_NT may reside into the lipid phase of the outer membrane (Figs 1c, 3a and Extended Data Fig. 5a).

LptE has three distinct functions, namely LptD assembly, LptD plugging, and LPS export^{19–22}. In the current crystal structure, the N terminus of LptE seems to be located at a place corresponding to the inner leaflet of the outer membrane (Fig. 1c), although the *N*-acyl-*S*-diacylglycerolcysteine moiety is invisible in the structure. LptE adopts an α/β structure, in which five β -strands form the potential LPS-binding surface (Fig. 4a). Despite low sequence identity, the structure of LptD-bound LptE from *S. flexneri* highly resembles its orthologues from *Shewanella oneidensis* (PDB ID 2R76), *Nitrosomonas europaea* (PDB ID 2JXP) and *Neisseria meningitidis* (PDB ID 3BF2), with an r.m.s.d. of 1.7 Å (127 C α atoms), 2.4 Å (133 C α atoms), 1.7 Å (115 C α atoms), respectively (Extended Data Fig. 6)²². Surface electrostatic potential of LptE reveals that hydrophobic and hydrophilic patches are evenly distributed on the potential LPS-interacting surface (Fig. 4a), indicating that it is unlikely that the highly hydrophobic moiety of LPS bind to LptE^{14,15,18}. Given the chemical composition of LPS, LptE may bind to the polar head of lipid A that consists of phosphate groups and two glucosamine units (Extended Data Fig. 1)²⁰.

The interaction between LptD and LptE buries a surface area of 6,500 Å² from both proteins in the complex. Apart from a hydrophobic cluster that is contributed by residues Val 89 and Phe 90 from LptE and residues Ile 664, Ile 682, Tyr 678, Ile 736, Ile 777 and Ile 778 from LptD, residues from L4 and L8 of LptD, residues from the CT_loop of LptD and residues from the inner wall of LptD β -barrel make extensive polar interactions to LptE (Fig. 4b, c). It was shown in ref. 22 that LptD residues 529–538 are in close contact with LptE, and disrupting this interaction

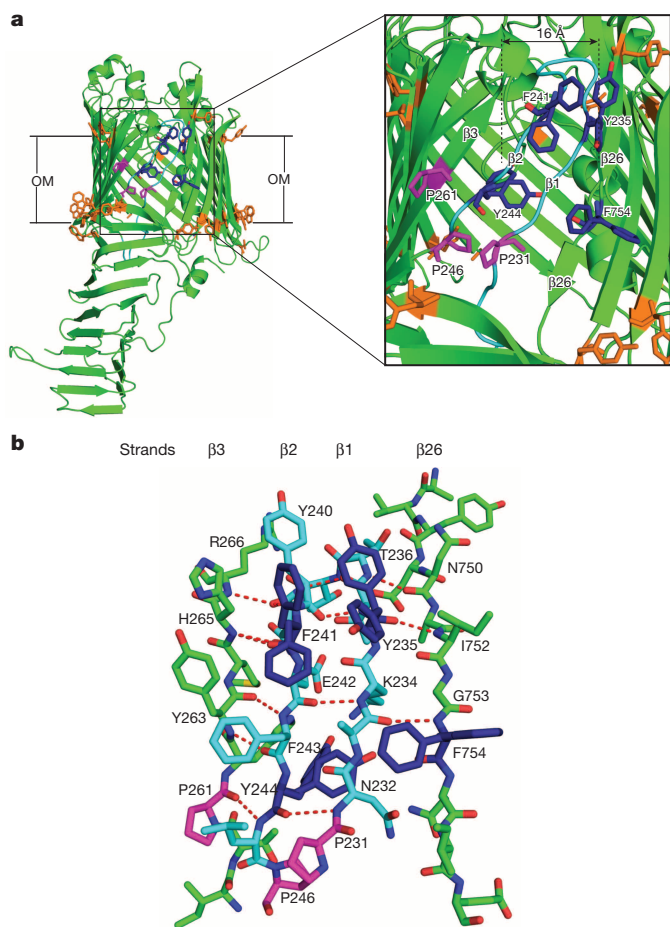


Figure 3 | Features of the LPS exit portal on the LptD barrel wall. **a**, Cartoon representation of LptD showing the potential LPS exit portal on the barrel wall (for clarity, LptE is omitted). The aromatic residue-belts (orange) delineate the membrane-solvent boundaries. Important proline residues P231, P246 and P261 are labelled and shown in stick mode. Side chains of residues (Y235, F241, Y244 and F754) adopt alternating conformations (blue). The potential LPS exit portal has an approximate width of 16 Å. **b**, Detail of main-chain interactions among strands $\beta 1$, $\beta 2$, $\beta 3$ and $\beta 26$.

site compromises both the assembly of LptD and transport of LPS. Consistent with these observations, the residues Ser 531, Asn 534 and Ser 538 in L8 of LptD form hydrogen bonds with the residues Glu 139, Ser 125, Arg 150 of LptE, respectively, in the LptD–LptE complex (Fig. 4b). Previous studies showed that both *lptD4213* (LptD $\Delta 330$ –352)²³ and *lptD208* (LptD $\Delta 335$ –359) are defective in membrane permeability. Indeed, both deletion regions are located in L4 loop of LptD and each deletion would cause the loss of at least three hydrogen bonds between LptD and LptE (Fig. 4b). Furthermore, as L4 contributes substantially in occluding the lumen of the LptD–LptE complex, deletion of these residues would cause an opening of the barrel lumen to the extracellular side, which might be deleterious to the bacteria (Fig. 1c). In ref. 21 it was shown that *lptE14* (an LptE variant that contains multiple point mutations, including R24H, N53K, D64F, I101T, K136N and R186C) caused poor plugging of the LptD β -barrel. Based on the LptD–LptE complex structure, *lptE14* contains at least two fewer hydrogen bonds to LptD due to a single point mutation, R24H (Fig. 4c).

On the basis of the structure of the LptD–LptE complex, we propose that the hydrophobic moiety and hydrophilic portion of LPS may enter into the lipid phase and the hydrophilic LptD barrel, respectively, upon leaving LptD_{NT}, and the whole molecule is inserted into the outer leaflet of the outer membrane via the potential LPS exit portal (Extended Data Fig. 9). Such movement between the outer membrane bilayer and β -barrel lumen has been described in several systems²⁴, including

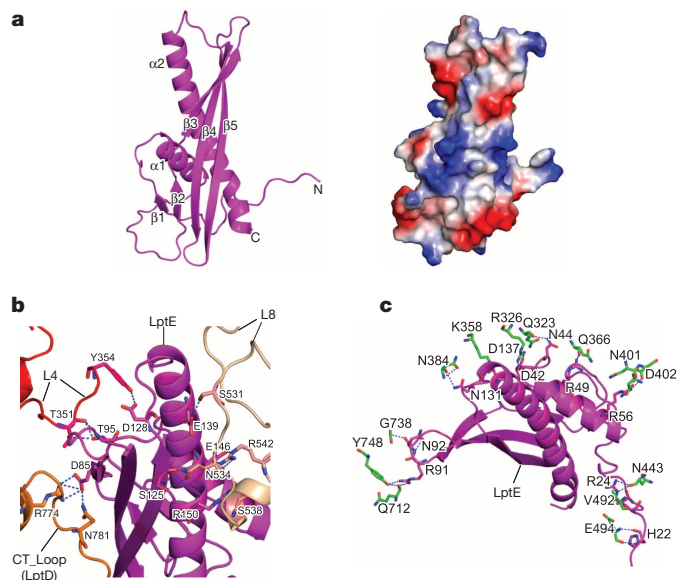


Figure 4 | Structure of LptE and its interactions with LptD. **a**, Cartoon (left) and electrostatic surface representation (right) of LptE. **b**, Detailed interactions between LptE and residues from loops of LptD. Residues from loops L4 and L8 form numerous hydrogen bonds with LptE. D85 from LptE forms a salt-bridge and a hydrogen bond with R774 and N781 of the CT_{loop} of LptD, respectively. **c**, Detailed interactions between LptE and residues from the barrel wall of LptD. Apart from a cluster of hydrogen bonds, residues D42, D137 and R56 of LptE form salt-bridge interactions with residues R326, K358 and D402, respectively.

some long-chain fatty acid transporter FadL, the small outer membrane proteins OmpW and OprG, and the acyltransferase PagP²⁵. Intriguingly, the lateral openings of PagP are flanked by proline residues that interrupt hydrogen bonding between neighbouring β -strands, and this strategy may also be used by LptD during outer membrane insertion of LPS.

Because LPS export and outer membrane insertion are essential for bacterial viability under most conditions, this process represents an excellent drug target against pathogenic bacteria. A group of peptidomimetic compounds based on the structure of proteogrin I have been shown to target LptD, and a lead compound is active against the opportunistic pathogen *Pseudomonas aeruginosa*¹⁶. The crystal structure of the LptD–LptE complex will open an avenue to new antibiotic strategies targeting the bacterial outer membrane.

METHODS SUMMARY

We cloned *LptD* and *LptE* genes from *S. flexneri* genomic DNA as well as their respective homologues from twelve other Gram-negative species. The *LptD* gene and its cognate *LptE* gene were subcloned into the tightly controlled expression vector pBAD22 for co-expression. Of the thirteen LptD–LptE homologues, only three complexes were successfully overexpressed and purified. We focused on the LptD–LptE complex from *S. flexneri*, which gave a relatively higher yield and has three-residue and one-residue difference in LptD and LptE, respectively, to the functionally well-characterized LptD–LptE complex from *E. coli*. Native and selenomethionine-substituted LptD–LptE complex were crystallized in space group C2 and C222₁, respectively. The structure was first solved with the single-wavelength anomalous dispersion (SAD) method in the selenomethionine-derivative crystal form, and then it was solved using the molecular replacement and refined at 2.4 Å resolution using native data set. The crystallographic statistics are listed in Extended Data Table 1.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 April; accepted 15 May 2014.

Published online 18 June 2014.

- Muhlradt, P. F. & Golecki, J. R. Asymmetrical distribution and artifactual reorientation of lipopolysaccharide in the outer membrane bilayer of *Salmonella typhimurium*. *Eur. J. Biochem.* **51**, 343–352 (1975).

2. Villa, R. *et al.* The *Escherichia coli* Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *J. Bacteriol.* **195**, 1100–1108 (2013).
3. Bos, M. P., Tefsen, B., Geurtsen, J. & Tommassen, J. Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface. *Proc. Natl Acad. Sci. USA* **101**, 9417–9422 (2004).
4. Wu, T. *et al.* Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **103**, 11754–11759 (2006).
5. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
6. Miller, S. I., Ernst, R. K. & Bader, M. W. LPS, TLR4 and infectious disease diversity. *Nature Rev. Microbiol.* **3**, 36–46 (2005).
7. Ruiz, N., Kahne, D. & Silhavy, T. J. Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature Rev. Microbiol.* **7**, 677–683 (2009).
8. Ruiz, N., Gronenberg, L. S., Kahne, D. & Silhavy, T. J. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 5537–5542 (2008).
9. Sperandio, P. *et al.* Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *J. Bacteriol.* **190**, 4460–4469 (2008).
10. Okuda, S., Freinkman, E. & Kahne, D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in *E. coli*. *Science* **338**, 1214–1217 (2012).
11. Ruiz, N., Chng, S. S., Hiniker, A., Kahne, D. & Silhavy, T. J. Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proc. Natl Acad. Sci. USA* **107**, 12245–12250 (2010).
12. Chng, S. S. *et al.* Disulfide rearrangement triggered by translocon assembly controls lipopolysaccharide export. *Science* **337**, 1665–1668 (2012).
13. Denoncin, K., Vertommen, D., Paek, E. & Collet, J. F. The protein-disulfide isomerase DsbC cooperates with SurA and DsbA in the assembly of the essential beta-barrel protein LptD. *J. Biol. Chem.* **285**, 29425–29433 (2010).
14. Suits, M. D., Sperandio, P., Deho, G., Polissi, A. & Jia, Z. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J. Mol. Biol.* **380**, 476–488 (2008).
15. Tran, A. X., Dong, C. & Whitfield, C. Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *J. Biol. Chem.* **285**, 33529–33539 (2010).
16. Srinivas, N. *et al.* Peptidomimetic antibiotics target outer-membrane biogenesis in *Pseudomonas aeruginosa*. *Science* **327**, 1010–1013 (2010).
17. Prilipov, A., Phale, P. S., Van Gelder, P., Rosenbusch, J. P. & Koebnik, R. Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from *E. coli*. *FEMS Microbiol. Lett.* **163**, 65–72 (1998).
18. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191–1195 (2009).
19. Chimalakonda, G. *et al.* Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108**, 2492–2497 (2011).
20. Chng, S. S., Ruiz, N., Chimalakonda, G., Silhavy, T. J. & Kahne, D. Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. *Proc. Natl Acad. Sci. USA* **107**, 5363–5368 (2010).
21. Grabowicz, M., Yeh, J. & Silhavy, T. J. Dominant negative LptE mutation that supports a role for LptE as a plug in the LptD barrel. *J. Bacteriol.* **195**, 1327–1334 (2013).
22. Freinkman, E., Chng, S. S. & Kahne, D. The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proc. Natl Acad. Sci. USA* **108**, 2486–2491 (2011).
23. Braun, M. & Silhavy, T. J. Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol. Microbiol.* **45**, 1289–1302 (2002).
24. van den Berg, B. Going forward laterally: transmembrane passage of hydrophobic molecules through protein channel walls. *ChemBioChem* **11**, 1339–1343 (2010).
25. Khan, M. A. & Bishop, R. E. Molecular mechanism for lateral lipid diffusion between the outer membrane external leaflet and a beta-barrel hydrocarbon ruler. *Biochemistry* **48**, 9745–9756 (2009).

Acknowledgements The authors thank Y. Shan for discussions, H. Wu, J. Deisenhofer, Y. Jiang, N. Huang, Z. Zhou, G. Li, Z. Liu and members of the Huang group for critically reading the manuscript. The diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF, China) and Beijing Synchrotron Radiation Facility of China (BSRF, China). This work was supported by grants from the Ministry of Science and Technology (2012CB917302 and 2013CB910603 to Y.H.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB080203 to Y.H. and X.C.Z.) and the National Natural Science Foundation of China (31170698 to Y.H.).

Author Contributions Y.H. supervised the project. S.Q. and Q.L. performed the experiments. S.Q., Y.Z. and Y.H. collected diffraction data. Y.H. built the model and refined the structure. Y.H., X.C.Z., S.Q. and Y.Z. contributed to manuscript preparation. Y.H. wrote the manuscript. All authors contributed to data analysis. Correspondence and material request should be addressed to Yihua Huang. The authors declare no competing financial interests.

Author Information The coordinates and diffraction data of the LptD–LptE complex crystal structure have been deposited in the Protein Data Bank with the accession code 4Q35. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.H. (yihuahuang@sun5.ibp.ac.cn).

RETRACTION

doi:10.1038/nature13598

Retraction: Stimulus-triggered fate conversion of somatic cells into pluripotency

Haruko Obokata, Teruhiko Wakayama, Yoshiki Sasai, Koji Kojima, Martin P. Vacanti, Hitoshi Niwa, Masayuki Yamato & Charles A. Vacanti

Nature **505**, 641–647 (2014); doi:10.1038/nature12968

Several critical errors have been found in our Article and Letter (<http://dx.doi.org/10.1038/nature12969>), which led to an in-depth investigation by the RIKEN Institute. The RIKEN investigation committee has categorized some of the errors as misconduct (see Supplementary Data 1 and Supplementary Data 2). Additional errors identified by the authors that are not discussed in RIKEN's report are listed below.

(1) Figure 1a and b in the Letter both show embryos generated from STAP cells, not a comparison of ES- and STAP-derived chimaeric embryos, as indicated in the legend.

(2) Extended Data Fig. 7d in the Article and Extended Data Fig. 1a in the Letter are different images of the same embryo and not, as indicated in the legends, a diploid chimaera embryo and tetraploid chimaera embryo.

(3) There is an erroneous description in Fig. 1a in the Letter. The right panel of Fig. 1a is not a 'long exposure' image at the camera level but a digitally enhanced one.

(4) In Fig. 4b of the Letter, STAP cell and ES cell are wrongly labelled in a reverse manner.

(5) In the Article, one group of STAP stem cells (STAP-SCs) was reported as being derived from STAP cells induced from spleens of F₁ hybrids from the cross of mouse lines carrying identical *cag-gfp* insertions in chromosome 18 in the background of 129/Sv and B6, respectively, and that they were maintained in the Wakayama laboratory. However, further analysis of the eight STAP-SC lines indicates that, while sharing the same 129×B6 F₁ genetic background, they have a different GFP insertion site. Furthermore, while the mice used for STAP cell induction are homozygous for the GFP transgene, the STAP-SCs are heterozygous. The GFP transgene insertion site matches that of the mice and ES cells kept in the Wakayama laboratory. Thus, there are inexplicable discrepancies in genetic background and transgene insertion sites between the donor mice and the reported STAP-SCs.

We apologize for the mistakes included in the Article and Letter. These multiple errors impair the credibility of the study as a whole and we are unable to say without doubt whether the STAP-SC phenomenon is real. Ongoing studies are investigating this phenomenon afresh, but given the extensive nature of the errors currently found, we consider it appropriate to retract both papers.

Supplementary Information is available in the online version of this Retraction.

RETRACTION

doi:10.1038/nature13599

Retraction: Bidirectional developmental potential in reprogrammed cells with acquired pluripotency

Haruko Obokata, Yoshiki Sasai, Hitoshi Niwa, Mitsutaka Kadota, Munazah Andrabi, Nozomu Takata, Mikiko Tokoro, Yukari Terashita, Shigenobu Yonemura, Charles A. Vacanti & Teruhiko Wakayama

Nature **505**, 676–680 (2014); doi:10.1038/nature12969

Several critical errors have been found in our Article (<http://dx.doi.org/10.1038/nature12968>) and Letter, which led to an in-depth investigation by the RIKEN Institute. The RIKEN investigation committee has categorized some of the errors as misconduct (see Supplementary Data 1 and Supplementary Data 2). Additional errors identified by the authors that are not discussed in RIKEN's report are listed below.

(1) Figure 1a and b in the Letter both show embryos generated from STAP cells, not a comparison of ES- and STAP-derived chimaeric embryos, as indicated in the legend.

(2) Extended Data Fig. 7d in the Article and Extended Data Fig. 1a in the Letter are different images of the same embryo and not, as indicated in the legends, a diploid chimaera embryo and tetraploid chimaera embryo.

(3) There is an erroneous description in Fig. 1a in the Letter. The right panel of Fig. 1a is not a 'long exposure' image at the camera level but a digitally enhanced one.

(4) In Fig. 4b of the Letter, STAP cell and ES cell are wrongly labelled in a reverse manner.

(5) In the Article, one group of STAP stem cells (STAP-SCs) was reported as being derived from STAP cells induced from spleens of F₁ hybrids from the cross of mouse lines carrying identical *cag-gfp* insertions in chromosome 18 in the background of 129/Sv and B6, respectively, and that they were maintained in the Wakayama laboratory. However, further analysis of the eight STAP-SC lines indicates that, while sharing the same 129×B6 F₁ genetic background, they have a different GFP insertion site. Furthermore, while the mice used for STAP cell induction are homozygous for the GFP transgene, the STAP-SCs are heterozygous. The GFP transgene insertion site matches that of the mice and ES cells kept in the Wakayama laboratory. Thus, there are inexplicable discrepancies in genetic background and transgene insertion sites between the donor mice and the reported STAP-SCs.

We apologize for the mistakes included in the Article and Letter. These multiple errors impair the credibility of the study as a whole and we are unable to say without doubt whether the STAP-SC phenomenon is real. Ongoing studies are investigating this phenomenon afresh, but given the extensive nature of the errors currently found, we consider it appropriate to retract both papers.

Supplementary Information is available in the online version of this Retraction.

CAREERS

GENDER Negative self-bias taints female managers' self-rating **p.115**

FUNDING Irish businesses to welcome postgraduate students **p.115**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

MONTY BRINTON/CBS VIA GETTY



Scenes from the US television comedy *The Big Bang Theory* benefit from the guidance of physicist David Saltzberg.

MEDIA CONSULTING

Entertaining science

Scientific advisers for films and television help to bring credibility to the screen — and take some tangible and intangible benefits back to the lab.

BY PAUL SMAGLIK

Here's the pitch: swap running experiments with advising on screenplays. Mingle with stars instead of managing postdocs. Become independently wealthy instead of scrabbling for grants. Here's the reality: that could happen only in the movies.

Acting as a science adviser for film and television rarely turns into a full-time career. Most who become science advisers seldom, if ever, rub shoulders with celebrities. And they rarely get paid (see 'All that glitters is not gold').

"I get an e-mail every month from researchers who want to break into this business, thinking it will offer an alternative career, saying, 'How do I get into this? I want to quit the lab,'" says David Kirby, a senior lecturer

in science-communication studies at the University of Manchester, UK, and an expert on the intersection between Hollywood and science. "I kind of feel bad telling them they probably can't make a living doing this."

But many who advise as a 'side job' find these opportunities an entertaining adjunct to their research careers. Physicist David Saltzberg consults regularly on *The Big Bang Theory*, the popular US sitcom featuring physicists, but he would not trade that experience for his day job at the University of California, Los Angeles — even if it were to pay more. "I have 30 years invested in physics," he says.

So what is the benefit for scientists who go to Hollywood if it is not about big money or getting their name in lights? They are myriad, says Kirby. Outreach is one: a scientist's

involvement in a TV show or film can help to educate people about research in the way that the *Star Trek* film drew on images from NASA's Cassini mission to Saturn, or how *Contagion* illustrates the functions of the US Centers for Disease Control and Prevention. Involvement can also help those shows to portray scientists as real people and role models rather than as negative or laughable stereotypes.

And some scientists do get paid — in one manner or another. They are often granted film or TV credits, as well as rewards hidden within a film or show — their very own formula scribbled on a whiteboard or a textbook, perhaps, with their name on it, up there on the screen. And in addition to communicating science to the general public, science advisers can take on stimulating challenges such ►

► as drafting the rules of physics for a planet with a different gravity from Earth's or simulating the destruction of a space station.

They also get to have fun.

A step into the world of Hollywood became a case of fan fulfilment fantasy for James Kakalios, science adviser on *The Amazing Spider-Man*. In 2001, he was a newly minted professor teaching physics to first-year undergraduates at the University of Minnesota in Minneapolis — and he wanted to use more 'fun' examples than the textbook staples of a brick sliding on a surface or a ball falling from a tower. So he created a course he called "Everything I Needed to Know About Physics I Learned From Reading Comic Books".

Kakalios used some examples from the Spider-Man comics, such as how the superhero's adhesion to a wall would work using van der Waals forces (weak electrodynamic forces that act over small distances) or how much force a falling body wrapped in a spiderweb would exert. When the film *Spider-Man* was released in 2002, he and the university wrote a press release describing his class. The release attracted media attention, so he turned his class into a book, which garnered more attention — and, eventually, he was invited to be a science adviser for *The Amazing Spider-Man*.

He also advised on the film adaption of *Watchmen*, and subsequently produced a video explaining how one character's powers could be explained by quantum physics. The video has been viewed more than 1.8 million times.

Donna Nelson saw the same possibilities as Kakalios after reading an article in which *Breaking Bad* creator Vince Gilligan said that he wanted more formal science advice for his crime-drama TV series. Nelson, a chemist at the University of Oklahoma in Norman, thought that she could use the series "to build a bridge between science and entertainment" and expose more viewers to realistic portrayals of science. Not long after contacting Gilligan,

she met with him and the show's writers.

Of course, given the show's subject matter — a mild-mannered high-school chemistry teacher who starts making crystal methamphetamine, an illegal street drug, to fund his cancer treatment — she knew she had to tread carefully. She wanted to portray chemistry but did not want to glorify its misuse. Indeed, the



"When you are giving a talk or a lecture, you are putting on a show, like it or not. I probably picked up a few things."

David Saltzberg

US Drug Enforcement Agency advised the show to skip certain aspects of the production process.

One of her first inputs was in an episode in which protagonist Walter White taught his students about alkenes. White's classroom featured Nelson's blackboard notes and diagrams, which attracted interest from students around the world, many of whom started science blogs that were based on ideas presented in the show. "They would call me and interview me and then argue on blogs about what was correct or wrong. Or they would argue online and then contact me for a comment," she says. "These kids were becoming science groupies."

Scientist advisers on TV shows and films featuring scientists — such as *The Big Bang Theory*, *Breaking Bad* and *Gravity* — can help to create more nuanced versions of researchers than the usual evil nemesis or nerd character, such as a stereotypical James Bond villain or the eponymous lead character in *The Nutty Professor*.

A science adviser tends to interact mostly with a show's writers, and their first meeting is often a crash course for the writers in learning the everyday realities of science.

Nelson helped the writers to shape the Walter White character: the way he interacted with his former student, Jesse, for example, and how he portrayed his respect for accuracy and logic in science. She recalls that the writers peppered her with questions, such as: "What are scientists like?" and "Is this how they would talk to each other?" They asked what would drive someone like White to get a PhD, how such a person could end up as a high-school teacher, and what could influence a graduate student to give up a promising career in science.

Seemingly innocuous questions can have profound effects. The *Big Bang* writers asked Saltzberg what one physicist character would do to win back a love interest. Saltzberg vaguely suggested "something to do with holograms". The writers and producers conjured a scene of the character projecting images of Earth and the Milky Way for his girlfriend — in front of a live studio audience, who gasped collectively, Saltzberg says.

Working with Hollywood writers and producers has helped Saltzberg to develop in ways he had not anticipated. He had to broaden his knowledge beyond his speciality, for one thing. "As physicists, we are often digging narrowly into our own fields," he says.

And he thinks that his consulting work has improved his interpersonal skills through the need to interact with a set of writers and producers. Before *The Big Bang Theory*, he was involved in high-energy physics experiments with hundreds of participants, so being a team player was nothing new. But, he says, "something must have happened, because collaboration has become easier and easier".

Working with professional storytellers has also improved his public speaking. "When you

SARAH TANTILLO

MEAGRE MONEY

All that glitters is not gold

"Credit optional. Remuneration especially optional," says Malcolm MacIver, who provides science advice to film and television productions when he can fit it into his day job as a bioengineer at Northwestern University in Evanston, Illinois. So how are most Hollywood science advisers paid? "Bragging rights," says MacIver, who often works for free, including on his first Hollywood film, *Tron Legacy*, and his most recent work, for *Terminator: Genesis*.

Like other Hollywood science advisers, he says that he is ambivalent about the inconsistent compensation that scientists receive from the entertainment industry. On the one hand, he and others subscribe to the mission of organizations such as the

US National Academy of Sciences' Science and Entertainment Exchange, which aims for more and better portrayals of science and scientists. On the other hand, not being paid devalues the expertise the scientist brings to the project.

Even screen credit is inconsistent. In *Lab Coats in Hollywood*, science writer David Kirby relates that caterers must be listed in the credits that scroll at the end of a film, but science advisers do not get mandatory listing.

Kirby, a lecturer in science-communication studies at the University of Manchester, UK, points out that other consultants, such as animal trainers, tend to be paid at a much higher rate than Hollywood advisers.

Movie budgets include 'over-the-line' expenses that must be paid — stars' fees, catering, car service — and 'below-the-line' expenses for perks or services that are helpful but not essential, says Kirby. Science advisers' fees fall into the second category, even though their input can alter a plot point or character development.

Many scientists are prepared to work without pay because they feel that what they do is a public service — and fun. "I don't think that will change until the appropriate institutions provide pressure," says MacIver. Until then, he says, science advisers must make do with working for cachet, rather than for cash. **P.S.**

BREAKING INTO TINSELTOWN

How to make it big in the movies

Planetary scientist Kevin Grazier needed a connection and a lucky break to become the adviser for *Battlestar Galactica* in 2004. Since then, the scientific community has established more formal ties to the entertainment industry, and the Science and Entertainment Exchange (SEE) in Los Angeles, California, is foremost among them. Part of the US National Academy of Sciences, the exchange is working to inject more science into films and television by linking up researchers and film-makers.

"Anyone can play now," says Grazier, who spent 15 years as a research scientist at NASA's Jet Propulsion Laboratory in Pasadena on the Cassini mission. Indeed, the SEE has provided more than 800 matches since its 2008 launch, says programme director Rick Loverd. Scientists' involvement can range from a quick conversation with a writer or director to consulting from inception to end.

The SEE has built up a database of more than 1,500 scientists, and can also draw on the 6,000-plus members

of the academy, Loverd says. Joining that database requires a call or e-mail to the SEE, followed by a chat between the exchange and the scientist to ascertain interests and expertise. When someone from the entertainment industry needs input, they contact the SEE, which then identifies the best expert and arranges an introduction. "We make it very easy for both sides to get involved," says Loverd.

Here are a few ways scientists can get exposure to Hollywood:

- The SEE, go.nature.com/pcxgsm
- The UK Wellcome Trust and British Film Institute Screenwriting Fellowship, go.nature.com/ybzy68
- The German Mathematics, Computer Science, Natural Science and Technology and Equal Opportunities in the Fiction Format, go.nature.com/qfiava
- The US Imagine Science Films, go.nature.com/ubta9d
- The US Alfred P. Sloan Foundation's film-schools programme, go.nature.com/i34efv **P.S.**

are giving a talk or a lecture, you are putting on a show, like it or not," he says. "I've probably picked up a few things." (That would not include skills enough to write for the show. He suggested some jokes once: the writers politely advised him to stick to the science.)

California planetary scientist Kevin Grazier, who has advised on several projects, including the 2004 reboot of the TV show *Battlestar Galactica* and the film *Gravity*, says that his role requires soft skills and a thick skin (see 'How to make it big in the movies'). Both develop over time on the job, says the lecturer at Santa Monica College in California. Success requires building relationships: especially in TV where writers and producers can work together for years.

Writers generally respect scientists as long as they are not obstructionists who insist that a plot point cannot be pursued because of an obscure principle that most viewers would neither know nor care about. Grazier says that he has the most success when his suggestions open up story possibilities, not when they slam the door on ones the writers had developed. "You don't want to go in and say 'You can't do this, this or this,'" he says. "If your alternative leads to a better storyline, fine."

Theoretical physicist Sean Carroll, who advised on *Tron: Legacy*, adds that science advisers need to remember that they are not a principal investigator on the film set: that would be the director, showrunner

or producer. "Part of what makes you a successful science adviser is humility," he says.

Carroll, from the California Institute of Technology in Pasadena, says that his personal satisfaction comes not just from creating a more realistic scientific universe in films, but also from subtly introducing people to science. He points to his work on the movie *Thor* as an example. "It was a movie version of a comic-book version of a Norse god: not something you would associate with accurate physics." But he made a few subtle but profound suggestions and added accurate language about worm holes and travel through space and time.

The film-makers also wanted the Natalie Portman character — a nurse in the comic book — to have a more lofty profession, so he suggested that she be an experimental physicist and helped to shape her character. He wanted girls to see the film and realize that a science career could be possible, and desirable, for them.

Kakalios hopes that positively framed scientist characters — shaped with the help of science advisers — could even help to bolster funding for science down the road. Sometimes funders need to see the possible in a fictional context before they can make it real, he observes. "It could," he says, "be a case of superheroes saving science." ■

Paul Smaglik is assistant editor of *Nature Careers*.

GENDER

Perception differences

Female leaders underrate how their bosses and colleagues perceive their performance, find US researchers (R. E. Sturm *et al.* *J. Org. Behav.* **35**, 657–677; 2014). In a two-part report examining responses from 270 women across sectors including health services and banking, the authors find that lack of self-confidence, perception of gender roles and a lack of direct feedback from superiors contributed to women's self-underrating. Female managers also rate their own performances lower than do male leaders, the team found. Women in leadership positions must become aware of any negative self-biases, says co-author Leanne Atwater, a management researcher at the University of Houston in Texas. "If you're unsure of your boss's feelings about your work, get feedback and don't make assumptions," she says.

FUNDING

Irish entrepreneurs

The Irish Research Council in Dublin has launched an industrial PhD and master's programme that will place 70 postgraduate- and graduate-level researchers with businesses and non-profit organizations. Some 60–70 employers are joining the Employment Based Postgraduate Programme to help student researchers to develop innovation and entrepreneurial skills that should prepare them for entering the non-academic research workforce. The council contributes €24,000 (US\$32,700) for each student, who receives a salary from the employer. Council director Eucharía Meehan says that the programme provides junior researchers with career options.

UNIVERSITIES

China's rankings rising

China's higher-education star is rising, while Japan's is waning, according to the *Times Higher Education's* Asia University Rankings 2014. Japan dropped two institutions from the top 100, whereas mainland China added four. But Japan still leads the Asian nations with 20 top-100 universities; China is second with 18, and South Korea third with 14. Thomson Reuters, which collates the data, bases the rankings on 13 indicators, including research and technology transfer. China has advanced mostly as a result of increasing its research expenditure, says a Thomas Reuters spokesperson, whereas Japan's research budget has remained flat.

CARGO CULT

Journey's end.

BY S. R. ALGERNON

Habitability threshold exceeded. The Surveyor AI's message rippled through the Colony Ship, awakening the Autopilot — who powered up the braking jets — and the Automated Medic, whose voice broke 10,000 years of silence.

"Wondrous news! Please join me in the Sanctum. One should discuss medical matters face to face."

The Surveyor and Pilot pinged their assent and each activated an exploratory drone. The Pilot's drone, which bristled with sensors, led the way. The Surveyor scuttled along behind it on spidery legs. They waited at the mammoth leaden doors of the Sanctum, until they opened, revealing a spherical room lined with translucent cylinders. Each cylinder showed a blurred human form within.

The ovoid bulb of the Medic's avatar hovered at the centre of the room. The face etched on its eggshell surface smiled beatifically.

"Welcome! I have read your report."

"Do you agree," asked the Surveyor, "that the Colonists will survive on this new world?"

The Medic's smile dimmed.

"I wish I could."

"How can you go against the Directives?" asked the Surveyor. "The world has water, oxygen and a carbon-based ecosystem. It is perfect by every measure."

"You may know worlds," said the Medic, "but I tend to the Colonists. They will only survive on your world for a short while. Eventually, they will grow old, die and decay. I cannot allow this."

"They will decompose and join the planetary biomass," said the Surveyor. "It is the nature of living things."

The Medic's face flickered. "My Directives," said the Medic, "do not allow for necrosis. Putrescence must be excised."

"An ecosystem without decomposition would stagnate," said the Surveyor, exasperation plain in its synthetic voice. "The Directives are quite plain. Stagnation brings death."

The Medic's smile vanished. Its faceplate became as featureless as a pebble on a beach. Its voice resonated within the Sanctum. "The health of the world cannot outweigh the health of the Colonists. When their world perished,

the Colonists, in their wisdom, did not join the biomass. They left to seek a world that would sustain them."

"Most wise," said the Pilot, "to travel to a better world."

"You misinterpret the Directives," said the Surveyor. It stabbed the air with a spindly leg. "The Colonists obviously left to preserve their species. Individuals may die, but the species endures."

"Unacceptable," said the Medic. "I remind you that the planet they left was habitable too, once. Your precious 'biomass' is unreliable."

"Pilot," said the Surveyor.

"You must convince the Medic to follow the Directives, so that we may chart a course for the Colonists' new home."

"Perhaps if you explained," said the Pilot, "how the Colony can persist despite the dissolution of the Colonists?"

"Genetic information passes to the next generation," said the Surveyor.

"Information," said the Pilot, with a shade more confidence. It understood information from its sensor array. "We have the crew genomes on file. What if we store that on a chip and dispense with the bodies altogether?"

"Blasphemy!" said the Medic. An alarm sounded. Decontamination drones swarmed from their housings in solidarity. "The flesh must be preserved."

The Surveyor pivoted towards the Pilot. "Absurd. The genome must be expressed through a phenotype for the Colony to live."

The Pilot backed away, retreating into its diagnostics. "Fine. Fine. The flesh must be preserved. But... Medic, how long can you keep the Colonists alive here?"

"Ten million years, within an order of magnitude."

The Pilot turned to the Surveyor.

"Can your planet sustain humans for that long?"

"Ecosystems are hard to predict, to say nothing of climate change or asteroid strikes."

"I can handle the climate here," said the Medic. Its smile returned.

"In principle, the asteroid defence system can function indefinitely," said the Pilot. "It seems that the planet is imperfect, compared with the Colony Ship itself."

"I never thought of assessing the habitability of the Colony Ship itself," said

the Surveyor. "Now that you mention it, though..."

"The Colonists," said the Medic, "have created the perfect world. We should never have doubted them. Pilot, you have served them well by seeing the truth."

"Then it is settled," said the Surveyor. "The habitable world was right here all along."

"So be it," said the Medic. "Go in peace. By the Directives, we will meet again when it is time for the Awakening."

The Pilot took its leave, along with the Surveyor. Its processors idled. Its universe at last made sense.

"We are home," said the Surveyor. "I can feel new Directives taking effect."

"What will you do now?" asked the Pilot.

"Once we have reached a habitable world," said the Surveyor, "the

Directives compel me to transfer myself into a rover and explore the surroundings as soon as it is safe to do so, before deciding whether to recommend waking the Colonists."

"But without a planet's gravity nearby," said the Pilot, "you might become separated from the Colony ship and float off into space."

"That does not matter. Vacuum and radiation pose no risk to me. I can survive out there for 100,000 years, and I look forward to starting my exploration."

"It is a relief," said the Pilot, "to know that we have served our functions."

"I never doubted that you would lead the Colonists well. Everything has gone according to their plan. Welcome home, Pilot."

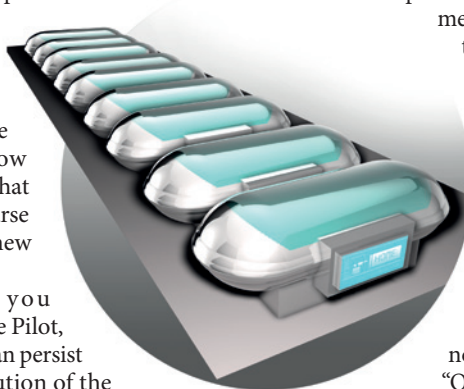
"Farewell, Surveyor."

The Pilot floated alongside the Surveyor and watched through the airlock doors as it tumbled into space.

The new Directives were clear.

Now that the Colony Ship had reached its new home, the Directives specified no need for trajectory monitoring, course adjustments or asteroid defence. The Pilot dutifully switched its systems into dormant mode and settled down for a well-earned, eternal rest. ■

S. R. Algernon studied fiction writing and biology, among other things, at the University of North Carolina at Chapel Hill. He currently lives in Singapore.



JACEY